ASME
SETTING THE STANDARD

Journal of Engineering for Gas Turbines and Power

Published Quarterly by ASME

VOLUME 129 • NUMBER 3 • JULY 2007

Transactions of the ASME, Journal of Engineering for Gas Turbines and Power (ISSN 0742-4795) is published quarterly (Jan., April, July, Oct.) by The American Society of Mechanical Engineers, Three Park Avenue, New York, NY 10016. Periodicals postage paid at New York, NY and additional mailing offices. POSTMASTER: Send address changes to Transactions of the ASME, Journal of Engineering for Gas Turbines and Power, c/o THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS, 22 Law Drive, Box 2300, Fairfield, NJ 07007-2300.
CHANGES OF ADDRESS must be received at Society headquarters seven weeks before they are to be effective. Please send old label and new address.
STATEMENT from By-Laws. The Society shall not be responsible for statements or opinions advanced in papers or ... printed in its publications (B7.1, par. 3).

TECHNICAL PAPERS

(Contents continued on inside back cover)

This journal is printed on acid-free paper, which exceeds the ANSI Z39.48-1992 specification for permanence of paper and library materials. ♾™
♻ 85% recycled content, including 10% post-consumer fibers.

(Contents continued)

Journal of Engineering for Gas Turbines and Power JULY 2007

Volume 129, Number 3

(Contents continued on facing page)

(Contents continued)

Journal of Engineering for Gas Turbines and Power JULY 2007

Volume 129, Number 3

TECHNICAL BRIEF

BOOK REVIEW

The ASME Journal of Engineering for Gas Turbines and Power is abstracted and indexed in the following:

AESIS (Australia's Geoscience, Minerals, & Petroleum Database), Applied Science & Technology Index, Aquatic Sciences and Fisheries Abstracts, Civil Engineering Abstracts, Compendex (The electronic equivalent of Engineering Index), Computer & Information Systems Abstracts, Corrosion Abstracts, Current Contents, Engineered Materials Abstracts, Engineering Index, Enviroline (The electronic equivalent of Environment Abstracts), Environment Abstracts, Environmental Science and Pollution Management, Fluidex, INSPEC, Mechanical & Transportation Engineering Abstracts, Mechanical Engineering Abstracts, METADEX (The electronic equivalent of Metals Abstracts and Alloys Index), Pollution Abstracts, Referativnyi Zhurnal, Science Citation Index, SciSearch (The electronic equivalent of Science Citation Index), Shock and Vibration Digest

Should Biomass be Used for Power Generation or Hydrogen Production?

Alessandro Corradetti

Umberto Desideri
e-mail: umberto.desideri@unipg.it

Dipartimento di Ingegneria Industriale,
Università di Perugia,
Via G. Duranti 93,
06125 Perugia, Italy

In the last several years, gasification has become an interesting option for biomass utilization because the produced gas can be used as a gaseous fuel in different applications or burned in a gas turbine for power generation with a high thermodynamic efficiency. In this paper, a technoeconomic analysis was carried out in order to evaluate performance and cost of biomass gasification systems integrated with two different types of plant, respectively, for hydrogen production and for power generation. An indirectly heated fluidized bed gasifier has been chosen for gas generation in both cases, and experimental data have been used to simulate the behavior of the gasifier. The hydrogen plant is characterized by the installation of a steam methane reformer and a shift reactor after the gas production and cleanup section; hydrogen is then purified in a pressure swing adsorption system. All these components have been modeled following typical operating conditions found in hydrogen plants. Simulations have been performed to optimize thermal interactions between the biomass gasification section and the gas processing. The power plant consists of a gas-steam combined cycle, with a three-pressure-levels bottoming cycle. A sensitivity analysis allowed to evaluate the economic convenience of the two plants as a function of the costs of the hydrogen and electrical energy.
[DOI: 10.1115/1.2718226]

Keywords: biomass gasification, hydrogen production, steam methane reforming, combined cycles

1 Introduction

Biomass is usually considered as a renewable energy source with low or zero life-cycle CO_2 emissions, since the carbon produced in the combustion can be balanced by the formation of new biomass (of course, this is true if the rate of biomass regeneration is equal to the rate of consumption). Therefore, in the future world energy scenario, biomass is expected to play a fundamental role for both greenhouse gas mitigation and traditional fossil-fuels replacement. According to the International Energy Agency (IEA) [1], biomass currently contributes as 10.8% of the world energy supplies (including waste). For clean power generation, biomass is generally employed in steam cycle and used directly as solid fuel in conventional "grate-type" boilers. The plant size is generally small (10–40 MW) due to economic reasons concerning the availability of biomass at low cost [2]. Therefore, such plants do not reach a high thermal efficiency (below 30%).

In the last several years, however, many researchers have focused their efforts on biomass gasification and pyrolysis. These processes offer the possibility of producing a gas rich in CO, H_2, and CH_4, which can be used as gas turbine fuel and generate power in a combined cycle, thus achieving the highest thermodynamic efficiency available on a commercial scale. Thermal efficiency of the biomass integrated gasification combined cycle (BIGCC) is slightly below 40% [3].

However, there is an additional opportunity for the utilization of biomass derived gas, which has recently gained a lot of interest: the gas can be treated in a downstream plant for hydrogen production. This option has been investigated by various researchers [4–6]. The idea of producing H_2 through biomass gasification is the natural answer to the increasing cost and depletion of natural gas and oil reserves, which are today the main sources of hydrogen, with over 75% of hydrogen in the world being produced by hydrocarbons processing [7]. In particular, steam methane reforming (SMR) is the most widespread technology for H_2 production. It is based on the reaction of hydrocarbons with steam, which generates hydrogen and carbon monoxide

$$C_nH_m + nH_2O \leftrightarrow \left(\frac{m}{2} + n\right)H_2 + nCO \tag{1}$$

Being that this reaction is endothermic, industrial steam reformers include a furnace where the tubes containing methane and steam are heated up to 800–900°C. Nickel is always used as catalyst to promote reforming reaction inside the tubes. The second important reaction that occurs in SMR is

$$CO + H_2O \leftrightarrow H_2 + CO_2 \tag{2}$$

This is known as the CO shift reaction because CO shifts to CO_2. Because it is exothermic, reactors operating at lower temperature are required to enhance H_2 production. Finally, H_2 is separated from CO_2 and other impurities with amine absorption or pressure swing adsorption.

This simplified scheme (reforming, cooling, shift, and separation) can also be used to treat the gas derived from coal or biomass gasification with currently available technologies. Because of the increase in depletion and cost of liquid and gaseous fossil fuels, solid fuels (such as coal and biomass) will constitute the better solutions for replacing traditional fossil-fuels in H_2 production in the medium term [8]. For instance, the U.S. has announced the creation of a zero-emission 275 MW power plant integrated with coal gasification, hydrogen production, and CO_2 capture and sequestration [9]. Therefore, biomass will be an important source of hydrogen in the medium term, especially for those countries without coal resources.

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received July 6, 2005; final manuscript received October 31, 2006. Review conducted by Dilip R Ballal. Paper presented at the ASME Turbo Expo 2005: Land, Sea and Air (GT2005), June 6–9, 2005, Reno, NV. Paper No. GT2005-69040

Several commercial or pilot plants based on biomass gasification have been built in the last few years, following different schemes and technologies, which can be grouped according to various principles: directly or indirectly heated gasifiers, air or oxygen blown gasifiers, and atmospheric or pressurized gasifiers. In a directly heated gasifier, combustion and gasification occurs in the same reactor. This is the case of the atmospheric pressure fluidized bed gasifiers, which are commercially operating with biomass. Pressurized directly heated gasifiers are still in the demonstration stage, though they are likely to be commercially ready soon. IGT (Institute of Gas Technology) gasifier is probably the most famous example representing this technology.

In an indirectly heated gasifier, the energy necessary for gasification reactions is supplied externally, so combustion occurs in a separate reactor. Heat is then transferred to the gasifier with exchange tubes, as in MTCI (Manufacturing and Technology Conversion International) gasifier, or by recirculating hot inert material, as in BCL (Battelle Columbus Laboratories) gasifier. In general, air-blown gasifiers are not suitable for hydrogen production because the syngas generated is very rich in nitrogen, and the equipment size and costs of the hydrogen plant increase. On the other hand, if oxygen is used as the oxidizer, the plant must be provided with an air-separation unit (ASU) to produce O_2, and ASU is a costly piece of equipment, too. An indirectly heated gasifier can avoid both these drawbacks, since air never mixes with syngas as the combustion occurs in a separate reactor.

The aim of the present paper is to assess the technoeconomic feasibility of a hydrogen plant and of a power plant based on the same gasifier. Therefore, the indirectly heated gasifier is the most appropriate technology. Other typologies of gasifiers, in fact, do not allow a comparison of the two plants on the same basis for the following reasons:

- Air-blown gasifiers are not suitable for the hydrogen plant, as previously explained.
- The installation of an ASU for gasifier feeding is not effective for the power plant, which can be suitably integrated with an air-blown gasifier [3].

Among indirectly heated gasifiers, the BCL gasifier was selected, even though it is still in the demonstrative phase, because of its high efficiency and the relatively high heating value of the gas produced. Furthermore, a lot of data concerning the performance of this gasifier are available in literature. ASPEN PLUS® has been used as a tool to evaluate gasifier and hydrogen plant performance, while the power plant has been modeled with GATECYCLE™. Finally, an economic analysis has been carried out to evaluate the economic feasibility of both plants.

2 Gasification Plant

The gasification plant has been modeled following the actual performance of the BCL gasifier, which was developed in the late 1970s at Battelle Columbus Laboratory. The Future Energy Resources Corporation (FERCO) has employed this gasifier in the SILVAGAS™ process, which is now under demonstration on a commercial scale at the McNeil Generating Station in Burlington, Vermont. The size of the plant is 182 dry tons per day of biomass feed [2]. The gas produced is burned in the boiler of the 50 MW power plant, which has a long history of operating with woody biomass, though conventional combustion technologies (a stoker grate) have been used in the past. In a future step, the gas will be utilized in a gas turbine and in the bottoming steam cycle, in order to improve the net efficiency of the plant.

The BCL gasifier is an indirectly heated gasifier, consisting of two main reactors: biomass is gasified only with steam, while combustion occurs in a second reactor, which is fed with the char generated by the first one and with air. The heat of combustion is transferred to the gasifier by recirculating hot inert material (usually sand) from the combustion reactor to the gasifier. This device

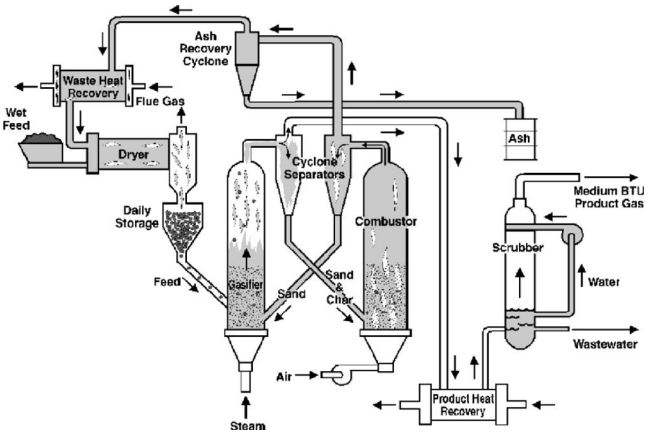


Fig. 1 SILVAGAS™ process [2]

allows the production of an N_2 free gas, without the installation of a costly air separation unit for O_2 production. Figure 1 illustrates BCL gasifier in the SILVAGAS process.

Figure 2 shows the ASPEN PLUS flow sheet of the gasification plant (note that dashed lines represent heat fluxes). The scheme is identical for power and H_2 production. The equilibrium reactors defined in the ASPEN PLUS library are not suitable for modeling the Battelle gasifier, since this reactor does not reach equilibrium conditions and the gas composition is deeply affected by the reactions kinetic. Therefore, yield reactors have been used for the simulation of the gasifier, together with a FORTRAN subroutine to control its yield. Real gasifier performance was taken from previous papers, concerning thermal and mass balance of the BCL gasifier [10].

Table 1 summarizes the inlet conditions of the streams entering the gasifier and the combustor. A typical wood composition

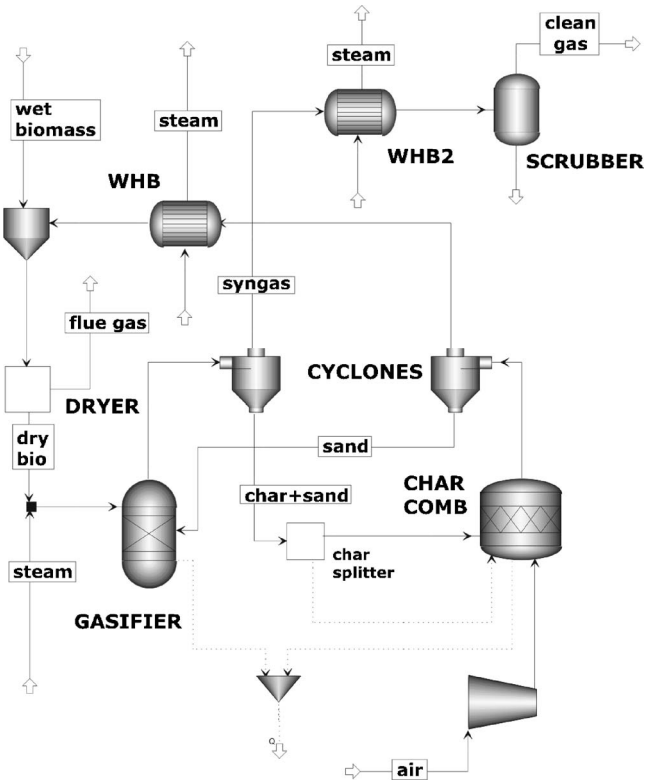


Fig. 2 ASPEN PLUS® flow sheet of gasification plant

Table 1 Assumptions for BCL gasifier simulation

Biomass composition	$CH_{1.54}O_{0.67}$
Ash content (ultimate %)	1.118
Wet biomass moisture (%)	30
Dry biomass moisture (%)	10
HHV dry biomass (MJ/kg)	19.46
Dry biomass input (t/hr)	54.1
Steam-to-biomass ratio (kg/kg dry)	0.019
Air-to-biomass ratio (kg/kg dry)	2
p gasifier (bar)	1.3

$CH_{1.54}O_{0.67}$ has been used in the simulation. The biomass daily input is 1856 tons per day on a wet basis (initial moisture is 30%), corresponding to 1300 dry tons per day (54.1 t/h). Dry biomass HHV is 19.46 MJ/kg; therefore, the equivalent thermal input is 292.4 MW.

Table 2 shows the gas composition and the overall performance of the gasifier, which can be expressed by a unique parameter, the cold gas efficiency. This is the ratio between the chemical energy of the produced gas and the feeding biomass energy content. According to the studies quoted before, the BCL gasifier has a cold gas efficiency of ~80%.[1] However, the simulation with ASPEN PLUS allows one to take into account all the variations in the performance of the gasifier. Once that the conditions and the composition of inlet streams have been fixed, there are just two parameters to specify in the simulation: the syngas composition and the cold gas efficiency. An ASPEN PLUS calculator block determines the char composition with a simple molecular balance of the species entering the gasifier. The combustion of this char in the CHAR-COMB provides the heat necessary to the gasifier. In general, if the thermal balance is not satisfied for a given gas composition and gasification temperature, then it is possible to adjust cold gas efficiency (thus changing syngas mass flow rate) in order to maintain the overall thermal balance. The model has been validated with the values presented by Breault and Morgan [10]: the thermal balance is satisfied if temperature, gas composition, and cold gas efficiency of this study are used in the simulation.

The size of the gasifier is assumed the same for hydrogen production and power generation. Therefore, the biomass daily input of 1300 dry tons per day has been determined on the basis of the heat rate required by the gas turbine chosen for power production, which has been modeled following the performance of a real machine, GE MS6001FA. Biomass is dried by the exhaust gases of the char combustor, after they have been cooled to 550°C in a waste heat boiler. This temperature has been determined to make the biomass moisture decrease to the required value of 10% in the dryer. Heat from syngas and flue gas is recovered in two boilers to generate saturated steam, which is sent to heat-recovery steam-generator (HRSG) superheaters in the power plant or used in the hydrogen plant. In the H_2 plant, steam pressure of the two boilers is 35 bar, corresponding to the SMR pressure. In the power plant, instead, boilers operate at different pressures, corresponding to the high and intermediate pressure of the HRSG.

For syngas cleaning, only the most significant component, from a thermodynamic point of view, has been modeled: the wet scrubber. The gas enters the scrubber at 253°C and exits at 60°C. The scrubber removes particulate, tars, and the sulfur and nitrogen compounds. The cooled gas is then compressed by a five-stage intercooled compressor (syn-COMP). Each stage has a polytropic efficiency of 0.75 and an interstage cooling temperature of 90°C. The gas pressure at the compressor outlet varies according to the plant to which it is sent. In the power plant a pressure of 20 bar is sufficient for gas turbine feeding, while in the hydrogen plant a pressure of 35 bar is required.

[1]Actually in the quoted reference [10], the syngas cold efficiency is defined as the ratio between the energy of the produced gas and all the energy inputs. However, in this case, the difference is negligible since the gasifier steam and power consumption is much lower than the biomass input thermal equivalent.

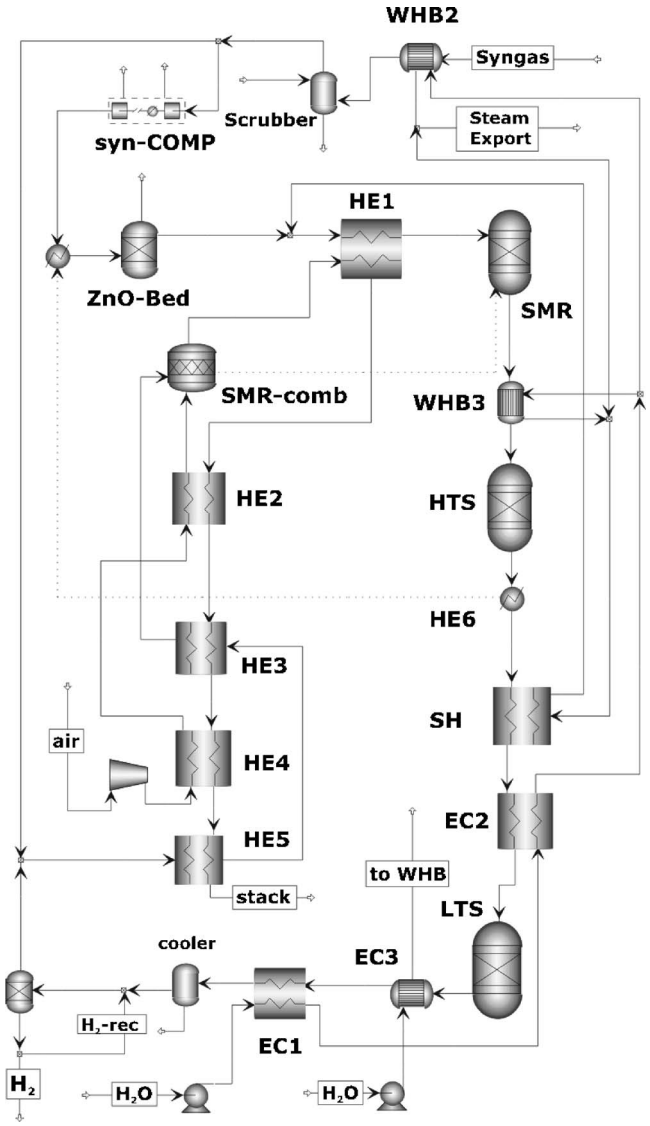


Fig. 3 ASPEN PLUS flow sheet of hydrogen plant

3 Hydrogen Plant

The clean gas, coming from the gasification and cleaning sections, has to be treated in a conventional plant for enhancing hy-

Table 2 Gasifier performance and syngas composition [10]

Gasifier outlet temperature (°C)		863
Combustor outlet temperature (°C)		970
Syngas yield (wet) (kg/ kg dry biomass)		0.99
Syngas wet composition (% vol)		
H_2O	19.90	
H_2	16.69	
CO	37.12	
CO_2	8.89	
CH_4	12.61	
C_2H_4	4.20	
C_2H_6	0.59	
Cold syngas efficiency (%)		78.5

Table 3 Composition of the main streams in H_2 plant

	COMP outlet	SMR outlet	HTS outlet	LTS outlet	PSA inlet	Off gas
T (°C)	179	900	413	200	40	38
p (bar)	35	34	32	31.1	30.5	1.3
Flow (kg/h)	40,647	134,436	134,436	134,436	70,695	66,063
Dry gas composition (% vol)						
H_2	20.86	60.04	64.55	65.39	65.43	26.08
CO	46.40	16.07	2.61	0.18	0.20	0.30
CO_2	11.05	23.26	32.28	33.88	33.87	72.61
CH_4	15.75	0.63	0.56	0.55	0.50	1.01
C_2H_6	0.73	0	0	0	0	0
C_2H_4	5.21	0	0	0	0	0

drogen production. The ASPEN PLUS flow sheet of the hydrogen plant is shown in Fig. 3.

Syngas undergoes four main treatments: reforming, cooling, CO shift, and purification. Among the technologies for steam reforming, we chose a conventional tubular steam methane reformer, in which the heat for the endothermic reforming reaction ($CH_4+H_2O=3H_2+CO$) is supplied with the heat generated by an external flame. The SMR temperature has been fixed at 900°C, a value commonly found in the commercial practice. Operating pressure is set by the compressor specification. Though high values of the SMR pressure decrease methane conversion, reformers usually work at pressure higher than 30 bar to reduce the size and cost of the equipment. Furthermore, high pressures increase the driving force for H_2 separation if a pressure swing adsorption (PSA) system is installed after syngas processing. Therefore, compressor discharge pressure has been set to 35 bar. SMR has been modeled with two reactors (Fig. 3): in the first (SMR), steam reforming of the syngas occurs, while in the second (SMR-comb) off-gas from PSA and a fraction of syngas are burned to meet the reforming heat demand. In particular, 12% of syngas is sent to SMR-comb.

The present analysis has been performed following the state of the art of the components; therefore, the plant layout is strongly influenced by common practice. For example, the syngas to be reformed is preheated in two steps: (i) before entering Zn-O beds the gas is heated up to 380°C by the stream leaving the high-temperature shift (HTS). (ii) It is then mixed with the process steam and heated in HE1 by the SMR exhaust gas to 650°C, which is a limit dictated by the resistance of materials. Zn-O beds are necessary to remove H_2S, which causes catalyst poisoning in SMR tubes.

As in commercial applications, the process steam is mainly produced in the boiler WHB3 (waste heat boiler 3) after the SMR, which cools the syngas down to 350°C. The preheating of water occurs in two economizers placed downstream the shift reactors. Additional steam at 35 bar (12% of the total) is needed to reach a steam to carbon ratio of 3.5 at SMR inlet. This steam is produced by the cooling of syngas coming from the gasifier in a second boiler (WHB2); this constitutes the main interaction between the hydrogen and the gasification plant. Saturated steam for export is produced both in the WHB2 and WHB, i.e., in the boiler fed by char combustor flue gas. The steam for reforming is super-heated to 365°C in SH and then mixed with syngas leaving ZnO-Beds.

The streams feeding SMR-comb are preheated by SMR exhaust gas: fresh air enters the combustor at 480°C, while the fuel is at 340°C. Methane conversion in SMR is 92%. Two CO shift reactors are placed downstream the reformer: the HTS is an adiabatic reactor operating at ~410°C and converts 82% of CO into CO_2; the low-temperature shift (LTS) reactor is an isothermal reactor in which CO conversion reaches 93%.

Table 3 shows the composition of the most significant streams. H_2 recovery in PSA has been set to 85% when H_2 concentration is 70% in the gas to be purified [4]. As the entering gas has a lower concentration, a fraction of pure H_2 is recycled back to obtain the required value. Thus, the actual recovery efficiency is 81.5%. H_2 is supposed to be produced at PSA feeding pressure (26 bar), since energetic and economic costs for H_2 compression (or liquefaction) and storage have not been included in the analysis (Table 4).

To characterize the overall performance of H_2 plants, it is possible to introduce the *fuel efficiency*, defined as the ratio between the energy content of the input feedstock and the energy content of the final hydrogen on a HHV basis. In the present study, the fuel efficiency of the H_2 plant is 78.7%, while the whole plant (including gasifier) has a fuel efficiency of 62.4% (note that fuel efficiency of the global plant can be determined by multiplying the cold gas efficiency of the gasifier by the fuel efficiency of the hydrogen plant). The plant achieves a reasonable fuel efficiency, which is lower than in SMR, but it is very high if compared to electrolysis based plants. Table 5 gives some values of fuel efficiencies for conventional H_2 plants.

Another important parameter to be evaluated is the *global thermal efficiency*, defined as the ratio of all the energy outputs to the inputs. Energy outputs must consider the enthalpy content of the exported steam, besides the chemical energy of H_2. On the other hand, energy inputs include all auxiliaries electricity consumption. With these assumptions, the global thermal efficiency is 72.9%.

Table 4 Energy balance of hydrogen plant

Biomass equivalent input HHV (MW)	292.56
Syngas compressor consumption (MW)	8.60
Auxiliaries consumption[a] (MW)	5.11
Steam export (t/h)	53.7
H_2 produced (kg/hr)	4,632
H_2 equivalent HHV (MW)	182.44
Fuel efficiency (% HHV)	62.4
Global thermal efficiency (% HHV)	72.9

[a]Includes electricity consumption of blower for gasification air, blower for SMR-COMB air, feed water pumps, gasifier auxiliaries (1% of biomass HHV), and pumps for heat rejection (from syngas compressor intercoolers and PSA gas coolers) equal to 1% of rejected heat.

Table 5 Fuel efficiency for various H_2 plants

H_2 production technology	Fuel efficiency (%)
Biomass gasification	62 %
Natural gas SMR[a]	75 %
Coal gasification[a]	58 %
Nuclear power plant+electrolysis[b]	21 %
Fossil-fuel-fired steam cycle+electrolysis[b]	34 %
Fossil-fuel-fired combined cycle+electrolysis[b]	49 %

[a]Reference [11].
[b]By assuming the following efficiencies: nuclear power plant=25%; steam cycle =40%; combined cycle=58%; electrolysis=85%; and grid loss=0%.

Table 6 Power plant assumptions and results

Gas turbine	
Model	GE PG6101 FA
Fuel	Syngas (natural gas)
Pressure ratio	15.6 (14.9)
Firing temperature (°C)	1,288
GT net power (MW)	73.87 (70.1)
Gas flow (kg/s)	13.05 (4.3)
Exhaust flow (kg/s)	205.3 (196.6)
Exhaust gas T (°C)	597
LHV efficiency (%)	34.66
Inlet air conditions	ISO
Syngas composition (% vol)	
H_2	19.56
CO	43.50
CO_2	10.41
CH_4	14.77
H_2O	6.14
C_2H_6	0.69
C_2H_4	4.92
LHV (kJ/kg)	16,325
Steam Cycle	
Pressure levels (bar)	97–20–4–0.05
Reheat T (°C)	538
Pinch point ΔT (°C) (HP, MP, LP evaporators)	20–15–10
HP steam from gasifier (% of total steam in HP-SH)	23.7
Approach point ΔT (°C) (MP, LP superheaters)	15–11
Steam turbine efficiencies (isentropic, mechanical)	0.9–0.94
MP steam from gasifier (% of total steam in MP-SH)	54.8
Stack T (°C)	89.4
Coolant flow (ΔT=+5°C) (m^3/h)	15,850
ST power output (MW)	52.30
CC power output (MW)	126.17
Fuel compressor power (MW)	8.10
Power plant auxiliaries[a] (MW)	1.23
Other auxiliaries[b] (MW)	3.94
BIGCC net power (MW)	112.90
BIGCC LHV efficiency (%)	45.1

[a]Includes electricity consumption of feed water pumps and pumps for heat rejection from condenser (1% of rejected heat).
[b]Includes electricity consumption of blower for gasification air, gasifier auxiliaries (1% of biomass HHV) and pumps for heat rejection from syngas compressor intercoolers (1% of rejected heat).

4 Power Plant

The selected gas turbine for the power plant is the GE PG6101FA. General Electric gas turbines have accumulated over 500,000 fired hours of syngas experience [12], and the FA class machine is a suitable technology for IGCC installation [13]. Because of the lower heating value of syngas, fuel mass flow is greater when burning syngas rather than natural gas, thus increasing the net power. For example, General Electric declares an increase from 70 to 90 MW for 6FA gas turbines fired by a typical composition syngas. Of course, the power increase depends on the fuel heating value. Among gasification technologies, the Battelle gasifier generates a syngas with a rather high heating value ($\sim$17 MJ/m^3), and, therefore, the increase in gas turbine (GT) power is limited.

In the present study, the fuel lower heating value (LHV) is approximately one-third of natural gas on a mass basis (Table 5). The new gas turbine power has been determined by a GATECYCLE simulation, using performance data for the GE 6FA available in the software library and assuming the same firing temperature as the natural-gas-fueled turbine (1288°C). The obtained value is 73.9 MW (natural-gas-fueled GT has a nominal power of 70.1 MW).

Table 7 H_2 and power plant capital costs

	H_2 plant (k$)	Power plant (k$)
Biomass gasification	27,360	27,360
Syngas cleaning	8210	8210
Syngas compressor	12,235	11,515
SMR	22,090	—
Shift reactors	11,200	—
PSA system	14,550	—
Hydrogen plant BOP	23,920	—
Gas turbine	—	18,700
Steam cycle	—	26,960
Power plant BOP	—	17,310
Plant process cost	119,570	110,050
Plant facilities	11,960	11,005
Engineering fees	17,935	16,510
Contingencies	26,300	18,160
Total capital costs	175,760	155,720

Exhaust gas feeds a heat recovery steam generator with three pressure levels and reheat. Following typical operating conditions of the S106FA GE combined cycles [14], maximum pressure levels have been set at 97 bar and maximum temperature at 538°C. The remaining pressure levels are 20 bar, 4 bar, and 0.05 bar for the condenser. Additional steam to the HRSG is generated in the gasification plant. In particular, 24% of the HP steam is generated in the boiler fed by char combustor flue gas and 55% of the MP steam is produced in the boiler fed by hot syngas. The net power generated by the steam cycle is 52.3 MW, giving a total CC power output of 126.2 MW. BIGCC net power is 112.9 MW for a global plant efficiency of 43.6% (LHV basis). Power plant data are summarized in Table 6.

5 Economic Analysis

Capital costs have been taken from literature and upscaled according to the power-law formula

$$C = C_0\left(\frac{s}{s_0}\right)^f \tag{3}$$

where s is the size of the equipment, C_0 and s_0 are the cost and the size of the reference equipment, and f is the scale factor.

Since the BCL gasifier is not commercially ready, no precise information concerning gasification plant cost is currently available. Therefore, the analysis was based on a recent publication about the SilvaGas process [2], where the total capital cost for the gasification plant was estimated to be $\$12\times10^6$ for a 400 tons per day dry biomass plant (excluding syngas cleanup and compression). Considering the size of the plant in the present study and a scale factor of 0.7, gasifier cost is $\sim\$27.4\times10^6$. This value is rather high if compared to previous studies [4], but it was assumed to include all the investments relative to gasification plant (biomass preparation and drying, piping, electronics, etc.). Syngas cleaning and cooling section cost has been fixed at 30% of the gasification plant. Capital costs for hydrogen plant equipment have been taken from Hamelinck and Faaij [6] and scaled up, according to (3), with a scale factor of 0.7. Power plant capital costs have been obtained from Lozza and Chiesa [15]. Plant facilities and engineering fees have been respectively considered as 10% and 15% of the total cost for process plant [3]. Table 7 reports capital costs for the two plants. The resulting specific capital cost for hydrogen plant is $\sim$\$38,000 for a capacity of 1 kg/h H_2. For power plant specific cost is \$1379 /kW installed.

For the economic analysis, a base case has been considered in which the cost of electricity (for H_2 plant auxiliaries) and biomass have been fixed, respectively, at \$0.05/kWh and \$20/t. Annual costs for the base case are shown in Table 8. For the hydrogen plant, they include biomass, electricity, and O&M (operation and maintenance) costs, while feedstock and O&M are the main cost

Table 8 H_2 and power plant operating costs

	H_2 plant	Power plant
Operating yearly hours	8000	8000
Biomass input (kt/yr)	618.55	618.55
Feedstock cost (MM $/yr)	12.371	12.371
Power input (GWh/yr)	109.6	—
Electricity cost (MM $/yr)	5.481	—
Steam export (kt/yr)	430	—
Steam (MM $/yr)	–2.148	—
O&M (MM $/yr)	8.788	7.786
Total operating costs (MM $/yr)	24.492	20.157

items for the power plant. O&M costs are assumed to be proportional to the investment (5% for both plants). Steam cost is assumed at \$5/t.

H_2 cost and electricity cost have been determined as the value which makes the net present value of the investment equal to zero. A plant life of 20 years has been assumed, and two different interest rates (10% and 15%) were used in the calculation. The assumption of so high IRs is justified by the two following considerations: First of all, the technology is very innovative; thus, the risks in investment are higher than in conventional fossil-fuels plants. Second, taxes have not been considered. For the base case with IR=15% the H_2 cost is \$11.8/GJ_{LHV}, while the electricity cost is \$0.050/kWh. If IR=10%, hydrogen cost is \$10.2/GJ_{LHV}, electric energy cost is \$0.043/kWh. If compared to conventional technologies for H_2 and power generation, these values demonstrate the economic feasibility of the biomass gasification in both cases. In fact, conventional steam methane reforming plant produces H_2 at \$5–8/GJ, while NGCC generates power at \$0.03–0.04/kWh. In Table 9, the cost of H_2 determined in this study is compared to the cost of H_2 produced with conventional technologies and to the cost of traditional fossil fuels.

Table 9 clearly points out that biomass derived H_2 will be competitive, not only with hydrogen produced by natural gas reforming, but also with other conventional fuels used for automotive applications. Furthermore, gasoline, diesel, and natural gas cost will reasonably increase with the depletion of fossil-fuels reservoirs, such as the cost of H_2 produced through natural gas reforming.

Figures 4(*a*) and 4(*b*) show the influence of biomass cost on H_2 and electricity cost for the chosen interest rates.

As an alternative method to assess the economical feasibility of the two plants, the pay-back period of the investment can be determined as a function of H_2 and electricity selling price. The results of this analysis is shown in Figs. 5(*a*) and 5(*b*). Three different biomass costs have been considered, and the interest rate has been fixed at 15%.

In order to evaluate the convenience of H_2 production over power generation, the current market price of both products should be considered. Unfortunately, the electric energy price is highly variable from one country to another, and it is not easy to fix the exact current selling price of hydrogen because, today, it is used more in industrial applications (e.g., ammonia synthesis and refining) than for transportation. It is also not easy to foresee when the infrastructure for transportation applications of hydrogen will be ready. The U.S. DOE has planned the transition to the marketplace between 2009 and 2025 (the commercialization decision by industry is expected for 2015), while the phase of expansion of markets and infrastructure is foreseen between 2015 and 2035. Anyway a key role for hydrogen competitiveness in the transportation field is the development of low-cost and more efficient technologies for its delivery, storage, and end use. For example, the U.S. DOE [8] has fixed the following milestones toward the hydrogen economy: (i) on-board hydrogen storage systems with a 9% capacity by weight to enable a 300 m driving range, (ii) hydrogen production from natural gas or liquid fuels at a price equivalent to \$1.50/gal of gasoline at the pump at 5000 psi; and (iii) hydrogen fuel delivery technologies that cost \$1.00/gal of gasoline equivalent.

Table 9 H_2 cost compared to other fuels cost

Fuel	Specific cost (\$/GJ)
H_2 biomass gasification (IR=15%)	11.8
H_2 biomass gasification (IR=10%)	10.2
H_2 conventional SMR[a]	5–8
H_2 electrolysis[a]	20–30
Natural gas[b]	3–4
Gasoline/diesel[b]	6–7

[a]Reference [16].
[b]Reference [17].

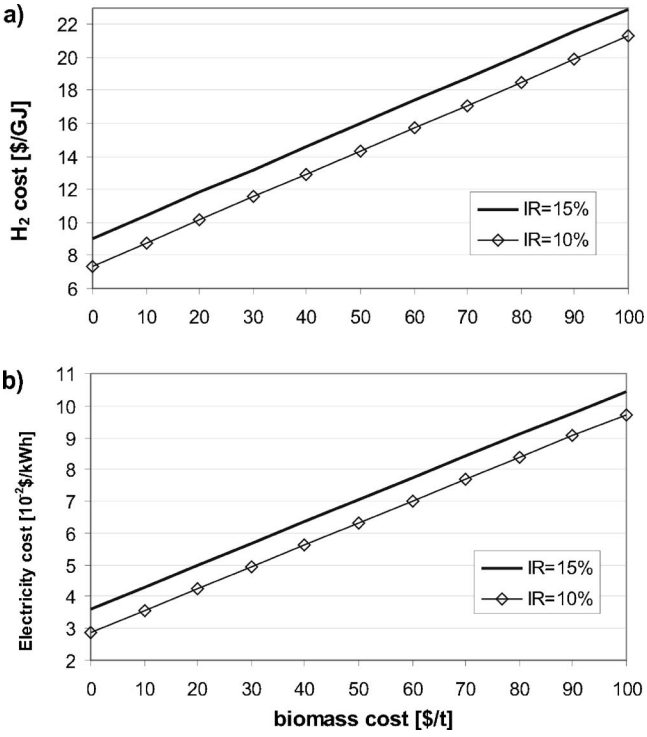


Fig. 4 Influence of biomass cost on H_2 and electricity cost (years=20)

Given all these uncertainties, a more general analysis has been carried out. The results are illustrated in Fig. 6, where the selling price of hydrogen, which would obtain the same economic revenue after 20 years with an interest rate of 15%, is shown for a given price of electric energy. The results for an IR of 10% are omitted as the obtained line is very close to the one illustrated (just shifted up of $\$9\times10^{-4}$/kWh). The dashed line shows an example of how to use the graph: if electric energy price is \$ 0.055/kWh a selling price of \$12.87/GJ is required to achieve the same goal in 20 years (corresponding to a net present value of \$ 29.1×10^{12} MM). In other words, the straight line may be considered to represent the breakeven between the two technologies.

Conclusions

A technoeconomical analysis has been carried out to investigate the feasibility of biomass gasification applied to power and hydrogen generation. An ASPEN PLUS simulation showed that hydrogen can be produced from wood with a fuel efficiency of 62%, while the efficiency of the BIGCC (biomass integrated gasification combined cycle) is ~44% (on LHV basis). Capital cost of the hydrogen plant is \$38,000/kg/*h* H_2. Capital cost of the power plant is \$

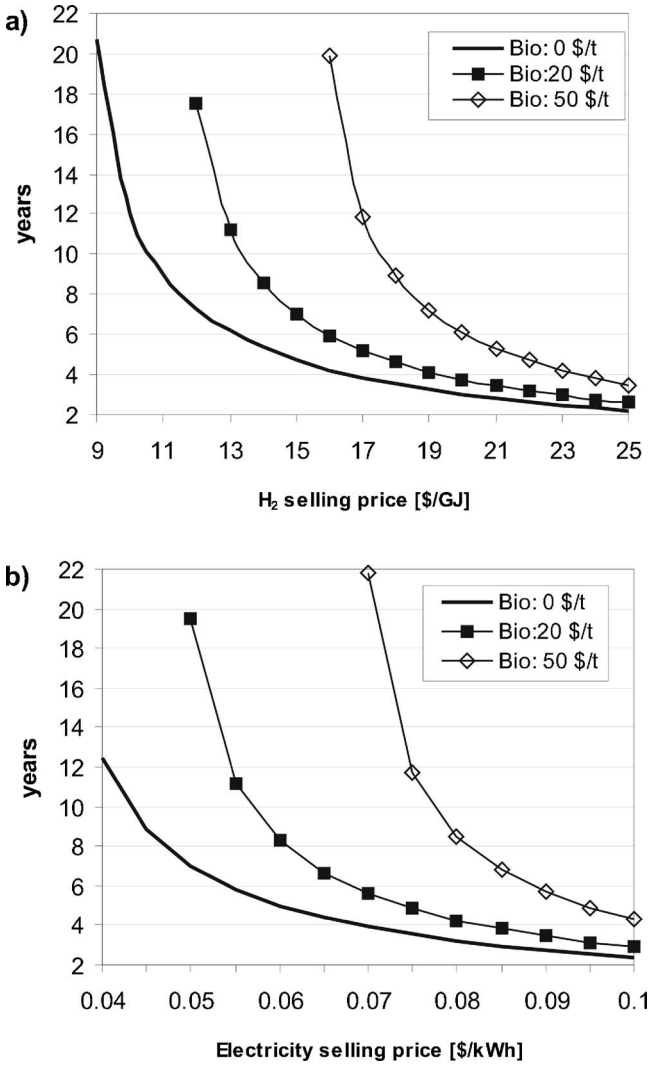


Fig. 5 Payback period of the investments over products selling price (IR=15%)

1379/kW installed. For a biomass cost equal to \$20/t, the resulting cost of produced hydrogen (\$11.8/GJ) is comparable to current technologies, such as those based on steam methane reforming. Also, the electric energy can be produced at a cost (\$ 0.050/kWh) competitive in the current market.

Therefore, even if much experience is needed in the integration of the gasifier with the remaining components of the plants, biomass gasification may represent in the medium term an important option to produce both power and hydrogen with high efficiency and competitive operating costs.

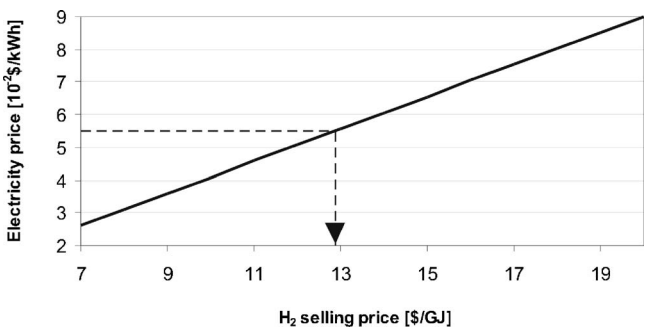


Fig. 6 Equivalence between the selling price of H_2 and electricity (NPV basis; IR=15%; 20 years; biomass cost=20\$/t).

Acknowledgment

The software ASPEN PLUS® Version 12.1 used in this paper is under license of Aspen Technology, Inc.

Nomenclature

ASU = air separation unit
BIGCC = biomass integrated gasification combined cycle
BOP = balance of plant
CC = combined cycle
COMB = combustor
COMP = compressor
EC = economizer
GT = gas turbine
HE = heat exchanger
HHV = higher heating value
HP = high pressure
HRSG = heat recovery steam generator
HTS = high-temperature shift
IR = interest rate
IGCC = integrated gasification combined cycle
LHV = lower heating value
LP = low pressure
LTS = low-temperature shift
MP = medium pressure
NPV = net present value
O&M = operations and maintenance
p = pressure
PSA = pressure swing adsorption
SH = super heater
SMR = steam methane reformer
syn = syngas
ST = steam turbine
T = temperature
WHB = waste heat boiler

References

[1] International Energy Agency (IEA), 2001, Web site, www.iea.org/Textbase/stats/PDF_graphs/29TPESPI.pdf
[2] Paisley, M. A., and Welch, M. J., 2003, "Biomass Gasification Combined Cycle Opportunities Using the Future Energy Silvagas® Gasifier Coupled to Alstom's Industrial Gas Turbines," ASME Paper No. GT2003-38294.
[3] Craig, K. R., and Mann, M. K., 1996, "Cost and Performance Analysis of Biomass-Based Integrated Gasification Combined-Cycle (BIGCC) Power Systems," NREL/TP-43021657 UC Category: 1311-DE96013105.
[4] Mann, M. K., 1995, "Techical and Economic Assessment of Producing Hydrogen by Reforming Syngas From the Battelle Indirectly Heated Biomass Gasifier," NREL/TP-431–8143 UC Category:1360-DE95009275.
[5] Williams, R. H., Larson, E. D., Katofsky, R. E., and Chen, J., 1995, "Methanol and Hydrogen From Biomass for Transportation," Energy for Sustainable Development, **1**(5), pp. 18–34.
[6] Hamelinck, C. N., and Faaij, A. P. C., 2001, "Future Prospects for Production of Methanol and Hydrogen From Biomass," Utrecht University, Copernicus Institute, Science Technology Society, Report No. NWS-E-2001-49, September.
[7] U.S. Department of Energy, 2002, "Energy Efficiency and Renewable Energy," Web site: www.eere.energy.gov, Vehicle and Fuel Cells, Fact of the Week, No. 205, Feb.
[8] U.S. Department of Energy, 2004, "Hydrogen Posture Plan—An Integrated Research, Development and Demonstration Plan," Feb., U.S. DOE, Washington, DC.
[9] U.S. Department of Energy, Office of Fossil Energy, 2004, "FutureGen—Integrated Hydrogen, Electric Power Production, and Carbon Sequestration Research Initiative," U.S. DOE, Washington, DC.
[10] Breault, R., and Morgan, D., 1992, "Design and Economics of Electricity Production from an Indirectly Heated Biomass Gasifier," for Battelle Memorial Institute by Tecogen, Inc., Waltham, MA, Oct. 22.

[11] Parsons Infrastructure & Technology Group, 2002, "Hydrogen Production Facilities Plant Performance and Cost Comparisons," Final Report, March, prepared for U.S. DOE National Energy Technology Laboratory.
[12] Shilling, N. Z., and Lee, D. T., 2003, "IGCC–Clean Power Generation Alternative For Solid Fuels," *PowerGen Asia 2003 Conference Proc.*, General Electric Company, Schenectady, pp. 1–9.
[13] Brdar, R. D., and Jones, R. M., 2000, "GE IGCC Technology and Experience With Advanced Gas Turbines," GE Power Systems, GER-4207 (10/00).
[14] Tomlinson, L. O., and McCullough, S., 1996, "Single-Shaft Combined-Cycle Power Generation System," GE Power Systems, GER-GER-3767C.
[15] Lozza, G., and Chiesa, P., 2002, "CO_2 Sequestration Techniques for IGCC and Natural Gas Power Plants: A Comparative Estimation of Their Thermodynamic and Economic Performance," *Proc. of International Conference on Clean Coal Technologies (CCT2002)*, Chia Laguna, Italy, Oct.
[16] Padró, C. E. G., and Putsche, V., 1999, "Survey of the Economics of Hydrogen Technologies," NREL/TP-570-27079, Sept.
[17] Haug, M., 2004, "Hydrogen—IEA Activities and Perspective," DTI: A UK/US Vision for Hydrogen Technology Conference, London, Oct. 11–12.

Simulation of Producer Gas Fired Power Plants with Inlet Fog Cooling and Steam Injection

Mun Roy Yap

Ting Wang

Energy Conversion and Conservation Center,
University of New Orleans,
New Orleans, LA 70148-2220

Biomass can be converted to energy via direct combustion or thermochemical conversion to liquid or gas fuels. This study focuses on burning producer gases derived from gasifying biomass wastes to produce power. Since the producer gases are usually of low calorific values (LCV), power plant performance under various operating conditions has not yet been proven. In this study, system performance calculations are conducted for 5 MWe *power plants. The power plants considered include simple gas turbine systems, steam turbine systems, combined cycle systems, and steam injection gas turbine systems using the producer gas with low calorific values at approximately 30% and 15% of the natural gas heating value (on a mass basis). The LCV fuels are shown to impose high compressor back pressure and produce increased power output due to increased fuel flow. Turbine nozzle throat area is adjusted to accommodate additional fuel flows to allow the compressor to operate within safety margin. The best performance occurs when the designed pressure ratio is maintained by widening nozzle openings, even though the turbine inlet pressure is reduced under this adjustment. Power augmentations under four different ambient conditions are calculated by employing gas turbine inlet fog cooling. Comparison between inlet fog cooling and steam injection using the same amount of water mass flow indicates that steam injection is less effective than inlet fog cooling in augmenting power output. Maximizing steam injection, at the expense of supplying the steam to the steam turbine, significantly reduces both the efficiency and the output power of the combined cycle. This study indicates that the performance of gas turbine and combined cycle systems fueled by the LCV fuels could be very different from the familiar behavior of natural gas fired systems. Care must be taken if on-shelf gas turbines are modified to burn LCV fuels.* [DOI: 10.1115/1.2718571]

Introduction

Since the discovery of fire, biomass has been used as a fuel and energy source. Approximately 14% of the world's population still uses biomass as an energy source [1], and for developing countries, biomass accounts for 35% of their fuel needs [1]. In the U.S., there are more than 350 biomass power plants producing about 7500 MW of electrical power, which is enough to power several million homes [2]. Utilizing proper management and advanced scientific methods of energy production, modern society has made biomass a sustainable and indigenous energy source.

Biomass is an attractive alternative fuel because it is renewable, sustainable, and indigenous. A common practice in agricultural, forest, and paper industries is the direct firing of biomass upon a boiler to create processed steam or for power generation purposes via steam turbines. This conventional method is subject to low efficiency and high emissions of NO_X, CO, and other pollutants. A typical small steam-boiler power system has an electric energy conversion efficiency ranging from 20% to 30%. A method to resolve the aforementioned challenges is to install a gas turbine (GT) between the combustion flame and the steam turbine. A modern gas turbine can withstand inlet gas temperatures at approximately 1430°C/2600°F. Exhaust gas exiting from the gas turbine can be utilized to boil and superheat steam via a heat recovery steam generator (HRSG). Combining the gas turbine and steam turbine systems forms the combined cycle system, which is used by many of the new natural gas fired power plants. A modern combined cycle efficiency ranges from 55% to 58%. A higher than 60% efficiency can be achieved by using an Advanced Turbine System [3].

Since direct firing of biomass fuel is unclean, gasification becomes a means of providing the gaseous fuel used for combustion in a gas turbine. Gasification is a partial combustion process, producing a composition of synthetic gas (syngas) or producer gas. The dominant reactant composition in the syngas is hydrogen (H_2) and carbon monoxide (CO). Producer gas, which is typically not as clean as the syngas, also contains traces of ethane, ethylene, or benzene, etc. The heating value of producer gas is low, typically around 5.5–7.5 MJ/Nm^3, approximately 15–20% of the heating value of natural gas. The producer gas needs to be cleaned to remove impurities like particles, tars, and other trace chemical elements before entering the gas turbine. In contrast to a conventional steam-boiler turbine system in which the exhaust gases are cleaned after the boiler or sometimes no cleaning is performed at all, in a gasification system the fuel is cleaned before entering the gas turbine system. This in turn results in a reduction in emissions, cost, and maintenance.

Biomass integrated combined cycle (BIGCC)—The combined cycle, flue-gas cleaning, and gasification are currently existing technologies. The main challenge is integrating these three technologies together to produce an affordable, highly efficient, environmentally friendly, and reliable system. New technology for using biomass in an integrated gasification combined cycle (BIGCC) is in development and is subject to continuous improvement. Some of the technical challenges require in-depth understanding of the fundamental mechanisms involved in the process to overcome them. Other challenges require development efforts in pulling together existing technologies to make it work. The ultimate challenge is to build demonstration plants, accumulate operating experience, and continuously improve the performance.

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received October 23, 2006; final manuscript received December 9, 2006. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2006: Land, Sea and Air (GT2006), May 8–11, 2006, Barcelona, Spain. Paper No. GT2006-90164.

Since the producer gases are usually of low calorific values (LCV) with less than 15% of the natural gas heating values, six to ten times more fuel mass flow is required to provide the adequate heating values to achieve rated power output. Therefore, additional fuel pump power will be needed, and the compressor backpressure will be increased accordingly. Under increased compressor pressure, the compressor stability margin will be reduced, and the decision needs to be made to either (a) maintain the same stability margin and pressure ratio by opening up the turbine nozzle throat area or (b) compromise the compressor stability margin (i.e., allow a narrower stability margin) but achieve a higher efficiency by operating the compressor with a higher pressure ratio. In the meantime, if a GT is fired with LCV fuels with an equal energy as NG, LCV fuels will reduce the turbine inlet temperature (TIT) due to increased sensible heats absorbed by the nonfuel gases in the gas flow (e.g., nitrogen and water vapor). Reduced TIT will result in reduced thermal efficiency. Furthermore, traditional approaches of using inlet air cooling or steam injection to augment GT output power could further strain the overloaded compressor when LCV fuels are used. It is not clear what would be the overall LCV-fired power plant performance under the influence of these factors at various operating conditions. This study is initiated to investigate potential issues and challenges of utilizing existing on-the-shelf gas turbines to burn the producer gases with the following ***objectives***:

- Assess and compare the performances of power plants including single steam turbine systems (Rankine cycle), single gas turbine systems (Brayton cycle), and combined cycle systems.
- Study the effects of ambient temperature and humidity upon the overall system performance.
- Study the different HRSG sizing upon the performance of the overall system.
- Access and compare performance of gas turbine inlet air cooling and steam injection (STIG) into combustion chamber.

The comprehensive analysis and assessment are documented in the report by Yap and Wang [4]. This paper reports part of the results.

Computational Tool and Simulated Cases

The commercial software THERMOFLOW (release 13, including THERMOFLEX, GT PRO, and STEAM PRO) is employed in this study. The input information includes GT model, ambient conditions, general power plant requirements and GT setup (inlet cooling, stem injection, water injection, etc), fuel type, optional combined cycle with related components, and economic criteria. During the simulation, the fuel is continuously added until the exhaust temperature reaches the designed value for the natural gas fired system. The simulation performs iteration by first calculating the combustion process using the given air mass flow rate and producing the result of the exhaust temperature. From the exhaust temperature and the combustion reaction, the compressor performance margin is checked with the existing GT model to ensure that the compressor can perform the required duty. If the compressor is not able to compress the needed amount of air, the iteration of the combustion is undertaken again with reduced fuel mass flow rate, and hence, lowering the turbine inlet temperature (TIT) and the pressure ratio (for LCV cases). The iteration follows the same method as described earlier until there is a convergence, which would suggest a possible actual performance of the GT with the given inputs. The above described default approach will use the GT exhaust temperature as a guideline. TIT can replace the exhaust temperature and be specified as the operating criterion. Details can be found in the THERMOFLOW manual [5].

Instead of designing a new turbine to burn LCV fuels, this study focuses on selecting an existing on-shelf commercial GT and examine its performance when LCV fuels are used. In this study, Rolls Royce 501KH5 GT is selected for the 5 MWe plant simulations. A companion study is also conducted for 20 MWe plants [4]. In that case, the Siemens GT 10 is selected.

Table 1 Fuel volume composition of the studied LCV producer gases

Volume composition	NG	PG1	PG2	PG3
H_2	0%	7.30%	11.62%	39.7%
H_2O	0%	23.00%	36.62%	0.00%
N_2	0%	37.20%	0.00%	0.00%
CO	0%	10.60%	16.82%	40.7%
CO_2	0%	14.60%	23.26%	19.6%
C_xH_y	100%	7.30%	11.68%	0.00%
HHV (MJ/kg) at 803 K	57.28	5.92	9.69	12.08
LHV (MJ/kg) at 803 K	51.62	5.02	8.23	11.16
HHV (MJ/kg) at 298 K	55.53	5.09	8.73	11.16
LHV (MJ/kg) at 298 K	50.05	4.32	7.40	10.31

The simulations are conducted under a controlled condition using the ISO condition (59°F, 60% relative humidity, and 1 atm) as the baseline case. Natural gas fired gas turbine is used as the reference case (Case 1). Producer gases of various calorific values are fed into the GTs designed for natural gas. The simulations are conducted first by feeding the producer gases without modifying the GTs but fixing the turbine inlet temperature (TIT) at the maximum allowable temperature. Then modifications are made by increasing TIT (Case 2b) or opening the first stage nominal nozzle area (Case 2c) in the turbine to reduce the compression backpressure. Case 2d is followed by optimizing both TIT and nozzle area. Comparisons are made among different cases. Typical producer gases consist of hydrogen (H_2), carbon monoxide (CO), hydrocarbons, carbon dioxide (CO_2), nitrogen (N_2), water (H_2O), tar, alkali, and other volatiles. For this study, three representative producer gases derived from biomass are used as shown in Table 1.

The compositions of first two LCV fuels are obtained from the Biomass Gasifier in Hawaii [6]. The lowest heating value 4.32 MJ/kg includes nitrogen while the medium heating value of 7.40 MJ/kg does not include nitrogen. The third producer gas, with a heating value of 10.31 MJ/kg, simulates a preheated dry syngas derived from an oxygen blown gasifier. The low heating value (LHV) of natural gas, 50.05 MJ/kg, serves as the reference. All of the fuels are assumed from the biomass gasifer delivered at a pressure of 10.34 bar and a temperature of 986°F (803 K) at inlet of the fuel compressor. The plant efficiency is calculated based on the fuel heating value at 298 K. A summary of all the cases simulated is shown in Table 2.

Results and Discussions

Under the ISO condition, 12 cases are simulated including: the simple cycle (gas turbine only), Rankine cycle (steam turbine only), and BIGCC (combined cycle with both gas and steam turbines). The three cycles are first simulated using natural gas followed by feeding different LCV fuels derived from biomass. The purpose of these simulations is to investigate the effects of different LHV fuels on the gas turbine and cycle performances.

Simple Gas Turbine Cases 1, 2a, 2b, 2c 2d, 3, and 4

Case 1: Natural Gas, ISO. Case 1 is the baseline reference case burning natural gas. In this case, natural gas is preheated to 803 K (986°F) to match the temperature of the producer gases delivered from the gasifier. The LHV is 51,624 kJ/kg and HHV is 57,282 kJ/kg at 803 K. One of the advantages of having a preheated fuel is to recover the low-grade waste heat to increase the overall thermal efficiency. In the simple cycle, the gas exhaust temperature is approximately 900 K. Figure 1 shows the graphical output from THERMOFLOW. The states of each point are given with "*p*" representing pressure (bar), "*T*" the temperature (degrees

Table 2 Summary of all simulated cases

Case	Cycle	Fuel	LHV (MJ/kg)	Ambient Temp (K)	RH (%)	Notes
1	Simple	NG	51.62	288.71	60	ISO GT NG
2a	Simple	PG1	5.02	288.71	60	ISO GT PG1
2b	Simple	PG1	5.02	288.71	60	TIT manipulation
2c	Simple	PG1	5.02	288.71	60	Nozzle manipulation
2d	Simple	PG1	5.02	288.71	60	TIT and nozzle manipulation
3	Simple	PG2	8.23	288.71	60	ISO GT PG2
4	Simple	PG3	11.16	288.71	60	ISO GT PG3
5	Steam	NG	51.62	288.71	60	ST ISO NG
6	Steam	PG1	5.02	288.71	60	ISO ST PG1
7	Steam	PG2	8.23	288.71	60	ISO ST PG2
8	Steam	PG3	11.16	288.71	60	ISO ST PG3
9	Combined	NG	51.62	288.71	60	ISO GT and ST NG
10	Combined	PG1	5.02	288.71	60	ISO GT and ST PG1
11	Combined	PG2	8.23	288.71	60	ISO GT and ST PG2
12	Combined	PG3	11.16	288.71	60	ISO GT and ST PG3
13	Combined	PG1	5.02	298.15	30	Ambient variation
14	Combined	PG1	5.02	298.15	90	Ambient variation
15	Combined	PG1	5.02	305.35	30	Ambient variaton
16	Combined	PG1	5.02	305.35	90	Ambient variation
17	Combined	PG1	5.02	298.15	30	Fog Cool Case 13 to 287.55 K 100% RH
18	Combined	PG1	5.02	298.15	90	Fog Cool Case 14 to 295.93 K, 100% RH
19	Combined	PG1	5.02	305.35	30	Fog Cool Case 15 to 293.13 K, 100% RH
20	Combined	PG1	5.02	305.35	90	Fog Cool Case 16 303.15 K 100% RH
21	Combined/ STIG	PG1	5.02	298.15	90	Max steam Case 14
22	Combined/ STIG	PG1	5.02	298.15	90	Same steam mass flow as fogger mass flow of Case 18
23	Combined/ STIG	PG1	5.02	305.35	30	Max steam Case 16
24	Combined/ STIG	PG1	5.02	305.35	30	Same steam mass flow as fogger mass flow of Case 19

Kelvin), and "*m*" the mass flow rate (kg/s). The pressure ratio of the GT is 11.3 and the TIT is 2150°F (1450 K). The net output power is 4.692 MWe and an efficiency of 30.32%. This case is used as the benchmark with which the other cases are compared in Table 3.

Case 2a: Producer Gas 1, ISO, without Modifying GT Nozzle Area. This is the first case using producer gas as the fuel. The firing of producer gas in the GT combustion chamber results in a higher mass flow rate and higher backpressure. The increased backpressure is attributed to two mechanisms: (a) the irrecoverable total pressure loss due to friction in the combustor and the turbine, and (b) the recoverable increase of stagnation pressure to push more mass flow through the choked nozzle. The pressure ratio increases to 13.4, and the TIT is lowered by 120.6°F (67 K) to 2029.73°F (1383 K). The increased pressure ratio may result in the need for a modified compressor or adding additional stages to handle an 18.6% increase of compression. The net power production is 5.322 MWe, which is 0.63 MWe (13.43%) greater than a natural gas fired GT under similar ambient conditions. The additional 13.43% of power increase is acceptable for the GT shaft and the generator as long as the increase is equal to or less than 25%. The reactants in the producer gas are composed primarily of CO and H_2. Combustion of these chemicals results in higher flame temperature, which may lead to more NO_x emission and may require higher combustion liner cooling. Since the fuel is of lower calorific value, a higher mass flow rate of producer gas is needed to reach the designed TIT value. The thermal efficiency for this system is 33.84%, about 3.52 percentage points (or 11.61%) higher than Case 1. The heat rate is correspondingly reduced by 10.4% from Case 1.

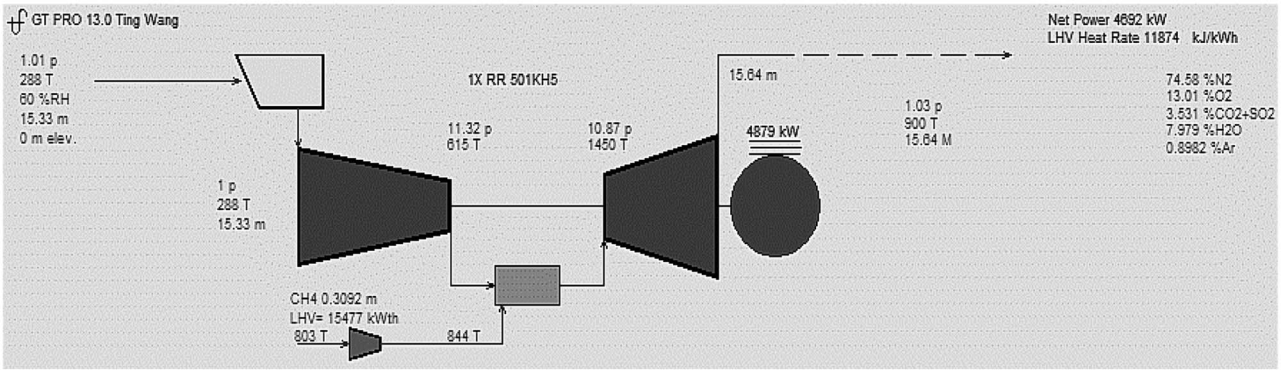


Fig. 1 THERMOFLOW GT graphical output of Case 1, where T (K), p (bar), and m (kg/s)

Table 3 GT simple cycle summary of Cases 1, 2a, 2b, 2c, 2d, 3, and 4[a]

Simple cycle GT	ISO	ISO GT Manipulation				ISO	ISO
Case number	1[b]	2a	2b	2c	2d	3	4
Ambient Temp (K) (59°F)	288.7	288.7	288.7	288.7	288.7	288.7	288.7
Rel. hum.	0.6	0.6	0.6	0.6	0.6	0.6	0.6
Pressure ratio	11.3	13.4	14.0	11.3	11.0	12.7	12.0
TIT (K)	1450	1383	1452	1421	1433	1451	1451
Air (kg/s)	15.3	15.3	15.3	15.3	15.3	15.3	15.3
Fuel LHV (MJ/kg)	50.05	4.32	4.32	4.32	4.32	7.40	10.31
Fuel HHV (MJ/kg)	55.54	5.10	5.10	5.10	5.10	8.73	11.16
Fuel (kg/s)	0.309	3.644	4.093	4.117	4.242	2.268	1.524
Fuel compressor power (MWe)	0.080	0.605	0.680	0.684	0.705	0.400	0.324
GT power (MWe)	4.879	6.050	6.848	7.006	7.171	5.913	5.291
GT LHV efficiency (%)	31.53	38.470	38.77	39.43	39.17	35.22	33.68
Net power (MWe)	4.692	5.322	6.039	6.192	6.335	5.394	4.854
LHV heat rate (kJ/kW-hr)	11874	10639	10529	10329	10403	11203	11651
LHV plant efficiency (%)	30.32	33.84	34.19	34.85	34.60	32.13	30.90
Net power increase (%)	0.00	13.43	28.71	31.97	35.02	14.96	3.45
Efficiency increase (%)	0.00	11.61	12.76	14.94	14.12	5.97	1.91
Heat rate increase (%)	0.00	−10.40	−11.33	−13.01	−12.39	−5.65	−1.88

[a]GT efficiency is based on fuel supplied at 10.34 bars and 803 K (986°F) from the gasifier. Net power=GT power - fuel compressor power- plant parasitic power (not shown). Plant efficiency represents the entire power plant efficiency based on fuel supplied at 298 K (77°F).
[b]Case 1 is the reference case for comparison with other cases.

The increased output power is contributed by increased gas flow rate and increased pressure ratio. The fuel compressor power is increased from 80 kW to 605 kW, which is consumed for compressing the producer gas to higher combustor pressure at higher fuel mass flow rate (3.644 kg/s producer gas compared to 0.309 kg/s of natural gas).

Case 2b: Producer Gas 1, ISO, Maximum TIT. Case 2a's TIT is 120.6°F (67 K) lower than the designed value. It is interesting to find out what the thermal efficiency would be if the TIT for producer gas fueled GT also reaches the designed value, 2150°F (1450 K). Therefore, in Case 2b, the TIT value is specified as 2150°F. When the option of the indirect control of TIT is selected, THERMOFLOW mimics the actual GT operation by using the turbine exhaust temperature to backcalculate the TIT. The iteration process gives a TIT of 1452 K instead of the assigned value of 1450 K. The results give a net power of 6.039 MWe with a compression ratio of 14.0 and thermal efficiency of 34.19%. In this case, the pressure ratio is increased, and the net power output is 28.7% higher than the natural gas fired Case 1. Replacement of a stronger GT shaft and a modified compressor unit may be necessary. There is a 3.87 percentage point (or 12.76%) increase in the efficiency while the heat rate decreases by 11.33%.

Case 2c: Producer Gas 1, ISO, Enlarged First-Stage Turbine Nozzle Area. Due to increased backpressure in both Cases 2a and 2b, the pressure ratios are higher than the designed value (11.3). With higher than the designed pressure ratio, the compressor is operated under a thin margin and instability is prone to be induced by surge or rolling stall. The partial-load performance can be questionable in Cases 2a and 2b, although the partial-load cases are not investigated in this study. To avoid the potential problems of operating the compressor too far away from the design point, in this case, the first-stage nozzles openings are enlarged by 23% to match the designed pressure ratio (11.3) as in Case 1. In addition, with the widening of the nozzle openings, the pressure ratio successfully reduces to 11.3, but the TIT is reduced to 2098°F (1421 K) while the net power increases 32% (6.192 MWe.) from Case 1. There is a need to replace the compressor, generator, and shaft due to the increased gross power. The plant thermal efficiency is rated at 34.85%, which is 4.53 percentage points (or 14.94%) higher than the natural gas fired Case 1.

Case 2d: Producer Gas 1, ISO, High TIT and an Enlarged First-Stage Turbine Nozzle Area. In this case, the TIT is assigned a value close to the maximum allowable value, and the first-stage nozzle openings are enlarged to reduce the same pressure ratio as in Case 1. After opening the first-stage turbine nozzles by 27.9% and increasing the TIT to 1433 K, the pressure ratio of 11.3 is maintained, and the net power output is 6.335 MWe with a thermal efficiency of 34.60%. Both the generator set and the turbine shaft may need to be replaced to accommodate the additional power produced.

Among the four Case 2's, enlarging the first-stage turbine nozzle area without raising TIT (Case 2c) gives the highest thermal efficiency (34.85%). However, opening the nozzle and raising the TIT to close to the maximum capacity (Case 2d) gives the highest output power (6.335 MWe), but the thermal efficiency drops 0.25 percentage point from Case 2c.

It must be noted that in the present study, although the overall plant thermal efficiency is calculated based on the fuel heating value at 298 K (77°F), the fuel compressor power is calculated based on the fuel supplied at 10.34 bar and 803 K. If the fuel compressor power is calculated based on the inlet condition at 298 K and 1 atm, it could require approximately 2.6 MW. If this power is entirely treated as a parasitic power, the net power output and thermal efficiency for all the cases burning PG fuels will need to be re-evaluated. Most likely, a NG-fired GT would have a better thermal efficiency than a PG-fired GT. In this study, the high fuel line pressure and temperature are treated as free energy from the gasifier.

Case 3: Producer Gas 2, ISO. The effects upon a GT with medium calorific value producer gas as a fuel source are studied in this case. The producer gas is rated as 8215 kJ/kg (LHV) and 9689 kJ/kg (HHV) and fed at 986°F (803 K). There is no adjustment to the TIT or the fist-stage nozzle openings. The results give a pressure ratio of 12.7 (12.4% higher than Case 1) with a TIT of 2152°F (1451 K). The net power production is 5.394 MWe, which is 15% higher than Case 1. Due to the higher fuel calorific value than in Case 2, the increased pressure ratio and net work output are acceptable to the selected GT. No specific modifications are needed. The thermal efficiency for this case is 32.13%, which is only 1.81 percentage points (or 5.97%) higher than Case 1 and 1.71 percentage points (or 5.1%) less than Case 2a. The fuel compressor power is 400 kW, which is lower than the fuel compressor consumption in Case 2a.

Case 4: Producer Gas 3, ISO. Producer gas 3 has the highest calorific value (11.156 kJ/kg LHV and 12.079 kJ/kg HHV) among the three producer gases selected in this study. The net power produced is 4.845 MWe, which is within the shaft power limits. This producer gas is actually a clean syngas with the com-

Table 4 Results of Cases 5–8 and comparisons of net electric power output and fuel mass flow rate with corresponding Cases 1, 2a, 3, and 4, respectively

Case	Net power (MWe)	Fuel mass flow rate (kg/s)	% power difference	% fuel flow rate difference
5	4.775	0.340	1.8	10
6	4.782	3.63	10.1	0.9
7	4.778	2.17	11.4	4.3
8	4.784	1.60	1.4	3.7

Table 5 Results of Cases 5–8 and comparisons of steam plant efficiency with corresponding gas turbine plant in Cases 1, 2a, 3, and 4, respectively

Case	η, steam cycle (%)	η, boiler (%)	% difference cycle efficiency
5	27.67	93.15	−8.7
6	30.54	91.12	−9.8
7	29.68	92.77	−7.6
8	29.29	93.81	−5.2

position purely of CO and H_2. The operating pressure ratio is 12, and the TIT is 2152°F (1451 K), both of which closely match the operating condition for natural gas fired Case 1. The plant thermal efficiency for this simulation is 30.90% and is marginally higher (1.9%) than Case 1.

In summary, the results in Table 1 show that fueling LCV producer gases to a simple gas turbine will require compressing more fuel mass flow to the combustor. The increased flue gas mass flow rate induces a higher backpressure and produces more output power. Even though the auxiliary power is required to compress a large amount of LCV fuel to a higher pressure, the overall net output power and net plant efficiency increase. The percentage of increased power and efficiency increases as the calorific value of the fuel reduces (see the increasing trend from Case 4, to Case 3 and to Case 2). More output power and higher plant efficiency can be further harnessed if the first-stage nozzle openings are adjusted 23–28% wider to match the pressure ratio of the designed value. Typically, a higher pressure ratio and higher TIT increase the Brayton cycle efficiency. However, in the cases studied here, lower pressure ratio and lower TIT can achieve similar or slightly better performance (see Cases 2c and 2d versus Cases 2a and 2b) because the higher pressure, induced by the LCV fuels, also increases more friction due to the significantly increased mass flow rate. In other words, the increased pressure ratio does not fully contribute to useful work.

Steam Turbine Only Plant: Cases 5, 6, 7, and 8

Case 5: Natural gas, steam turbine, ISO

Cases 6, 7, and 8: Producer gases, steam turbine, ISO. Using the ISO condition, a steam cycle is simulated with a boiler fired with natural gas. The steam turbine cycle is designed to produce roughly 5 MWe. The purpose of this case is to compare the performance differences between the stand-alone GT and the steam turbine (ST) with various fuels. The adiabatic flame temperature of natural gas combustion is greater than 2000°F (1366.5 K) in the boiler with a 93.14% thermal efficiency. The net power produced in Case 5 is 4.775 MWe with a steam cycle efficiency of 27.67%, which is about 2.65 percentage points (8.74%) lower than the GT performance in Case 1 (Table 4).

The same LCV producer gases used in Cases 2, 3, and 4, are used in Cases 5, 6, and 7, respectively, in a steam power plant. In general, a sugarcane or wood mill needs steam for drying or for other manufacturing processes, so most plants produce steam in house by burning sugar cane bagasse or wood wastes in the boiler. In these simulated cases, no process steam is tapped, and all the steam is used for power generation. The inlet and exit condition of the steam turbine values are assigned as a standard design value regardless of the type of fuel used, and hence, the fuel is fed and fired until the designated superheated pressure (42.4 bar) and temperature (644 K/700°F) are reached.

Tables 4 and 5 summarize the results of stand-alone steam turbine performance for Cases 5, 6, 7, and 8 and their comparisons with the corresponding GT Cases 1, 2a, 3, and 4, respectively.

In the GT plant, the parasitic power for compressing fuels is considered, but in the steam turbine plant, the fuel transport power is small and not considered because the furnace is operated at less than 2 atm pressure. For the steam cycle cases, the effect of calorific value on the steam cycle performance is not as pronounced as in the single GT simple cycle cases. Producer gases of all three heating values render almost identical thermal efficiency at 30.54%, 29.68%, and 29.29%, respectively, from lower heating value (4.32 MJ/kg) to higher heating value (10.31 MJ/kg). These efficiencies are all higher than the natural gas fired boiler in Case 5.

Combined Cycles with Various fuels: Cases 9–12

Case 9: Natural Gas, Combined Cycle, ISO

Cases 10, 11, and 12: Combined Cycle, ISO, Various Producer Gases. Case 9 simulates a combined cycle with the configuration of a GT, a heat recovery steam generator (HRSG), and a condensing nonreheat ST. Figure 2 shows the schematic output of Case 9, where each important state is shown next to the associated line of flow. Natural gas is fired in the same GT at ISO conditions as in Case 1, which provides a benchmark reference for the combined cycle systems.

In Case 1, with the GT alone, thermal efficiency achieves 30.32%. With the given combined cycle, heat is recovered via the HRSG. The superheated steam is fed into the ST to produce more power; hence the output power and the thermal efficiency are both significantly increased. The natural gas flows at 0.31 kg/s into the combustion chamber. In the combined cycle, the GT performance is identical to the cases where the GT is used as a simple cycle. The GT provides approximately 69.6% of the total power output, while the ST provides the rest of the power. Compared with the simple GT cycle of Case 1, the net output power of the combined cycle (Case 9 in Table 6) increases 22.7%, and the plant efficiency increases 14.72 percentage points (or 48.55%).

The effects of LCV fuels to the combined cycle are similar to those on the simple GT system (efficiency increases 2–8% and output power increases 2–13%) because GT produces about 70% of the total plant power, and the effect of LCV fuels to the steam turbine is limited as shown in previous steam turbine Cases 6, 7, and 8. The percentage of increased power and efficiency decreases as the caloric value of the fuel increases (see the descending trend from Case 10, Case 11 to Case 12 in Table 6).

Combined Cycles Under Different Ambient Conditions

Cases 13–16: Producer Gas 1, Combined Cycle, Various Ambient Conditions. The performance of a combined cycle burning producer gas at various ambient conditions is simulated in cases 13–16. Four representative ambient conditions are employed to examine how LCV fuels respond to ambient weather changes:

- Medium temperature and low humidity (77°F and 30% RH)

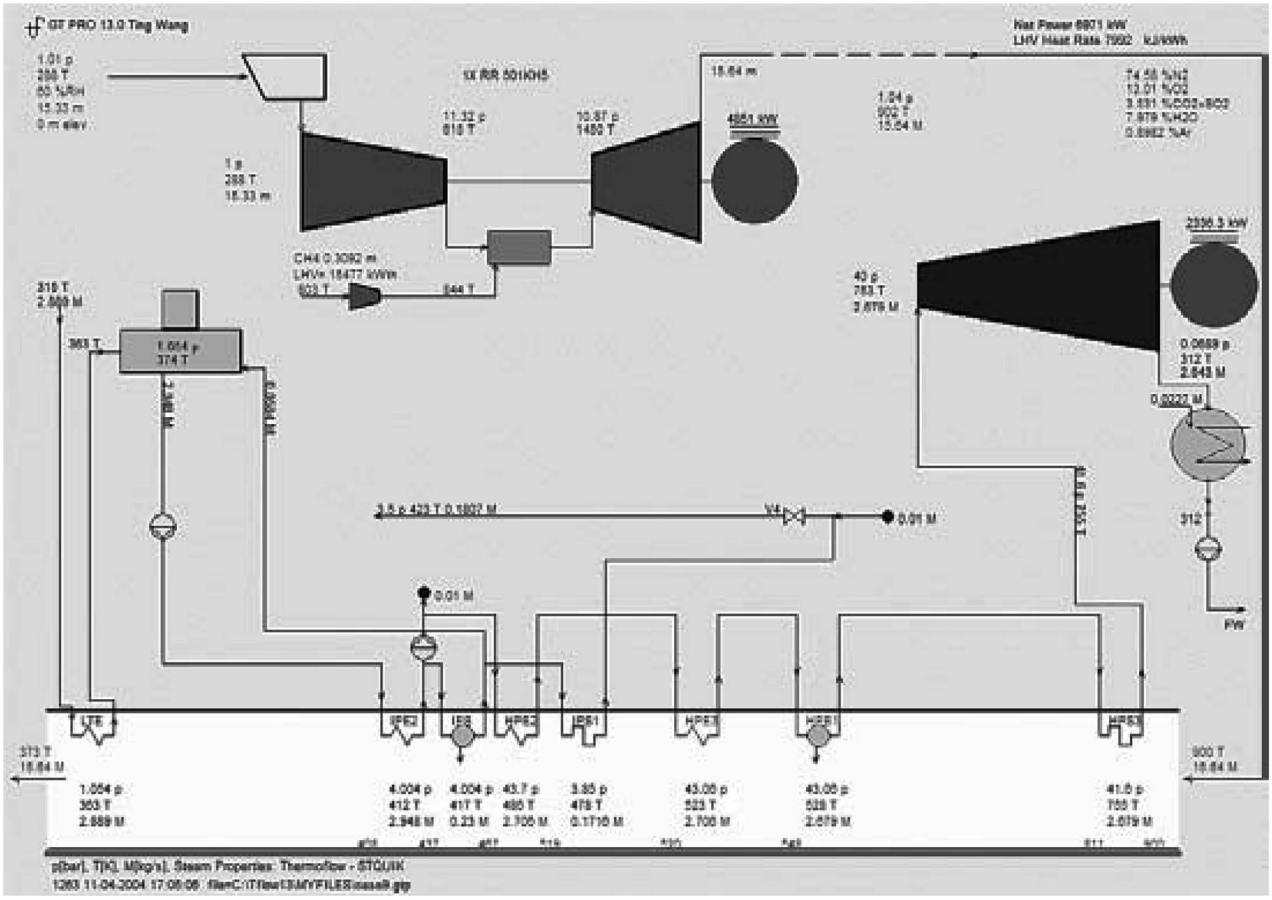


Fig. 2 THERMOFLOW Case 9 schematic output

- Medium temperature and high humidity (77°F and 90% RH)
- High temperature and low humidity (90°F and 30% RH)
- High temperature and high humidity (90°F and 90% RH)

Cases 9–12 are simulated at ISO condition while Cases 13–16 are simulated with four different ambient conditions as shown in Table 7. The fuel source is the producer gas 1 with 5.015 MJ/kg LHV supplied at the temperature of 986°F (803 K). Comparing

Table 6 Combined cycle performance for Cases 9, 10, 11, and 12 at ISO condition

Combined cycle	ISO			
Case number	9[d]	10	11	12
Ambient Temp, K (590°F)	288.2	288.2	288.2	288.2
Rel. hum.	0.6	0.6	0.6	0.6
Pressure ratio	11.3	13.4	12.4	11.9
TIT (K)	1450	1385	1408	1431
Air (kg/s)	15.3	15.3	15.3	15.3
Fuel LHV (MJ/kg)	50.05	4.32	7.40	10.31
Fuel HHV (MJ/kg)	55.54	5.09	8.73	11.16
Fuel (kg/s)	0.309	3.659	2.123	1.483
Fuel compressor power (MWe)	0.080	0.608	0.375	0.315
GT power (MWe)	4.851	6.067	5.508	5.133
ST power (MWe)	2.336	2.555	2.526	2.422
Net total power (MWe)	6.971	7.870	7.518	7.102
LHV GT heat rate (kJ/kW h)	11,486	9371	10,278	10,728
Total LHV heat rate (kJ/kW h)	7992	7224	7530	7754
LHV GT efficiency (%)	31.34	38.42	35.03	33.56
Total LHV thermal efficiency (%)	45.04	49.84	47.81	46.43
GT power increase (%)	0.00	25.07	13.54	5.81
Total power increase (%)	0.00	12.90	7.85	1.88
GT efficiency increase (%)	0.00	22.59	11.77	7.08
Total LHV efficiency increase (%)	0.00	10.66	6.15	3.09
Total LHV heat rate increase (%)	0.00	−9.61	−5.78	−2.98

[c]GT efficiency is based on fuel supplied at 10.34 bars and 803 K (986°F). Net power=(GT+ST) power - fuel compressor power- plant parasitic power (not shown). Total efficiency is based on fuel supplied at 298 K (77°F).
[d]Case 9 is the reference case for comparisons with other cases.

Table 7 Results of different ambient conditions for combined cycle Cases 13–16 using producer gas 1

Combined cycle	ISO	Various ambient conditions			
Case number	10[e]	13	14	15	16
Ambient temp (K°(F)	288.1 (59)	298.1 (77)	298.1 (77)	305.3 (90)	305.3 (90)
Rel. hum.	0.6	0.3	0.9	0.3	0.9
Pressure ratio	13.4	12.9	12.8	12.5	12.5
TIT (K)	1385	1398	1397	1407	1404
Air (kg/s)	15.3	14.6	14.5	14.2	14.0
Fuel LHV (MJ/kg)	4.32	4.32	4.32	4.32	4.32
Fuel HHV (MJ/kg)	5.09	5.09	5.09	5.09	5.09
Fuel (kg/s)	3.659	3.535	3.551	3.452	3.475
Fuel compressor power (MWe)	0.608	0.587	0.590	0.573	0.577
GT power (MWe)	6.067	5.859	5.881	5.718	5.749
ST power (MWe)	2.555	2.534	2.553	2.521	2.548
Net total power (MWe)	7.870	7.664	7.701	7.524	7.577
LHV GT heat rate (kJ/kW h)	9371	9372	9381	9377	9391
Total LHV heat rate (kJ/kW h)	7224	7165	7163	7127	7125
LHV GT efficiency (%)	38.42	38.41	38.38	38.39	38.33
Total LHV thermal efficiency (%)	49.84	50.25	50.26	50.51	50.53
GT power increase (%)		−3.43	−3.07	−5.75	−5.24
Total power increase (%)		−2.62	−2.15	−4.40	−3.72
GT efficiency increase (%)		−0.03	−0.08	−0.05	−0.21
Total LHV efficiency increase (%)		0.82	0.84	1.34	1.38
Total LHV heat rate increase (%)		−0.82	−0.84	−1.34	−1.37

[e]Case 10 is the reference for comparisons with other cases. See other notes in Table 6.

the results of Cases 13, 14, 15, and 16 with ISO Case 10, the output power of GT decreases by a range of 3.1–5.7%, while the plant total power is reduced by 2.2–4.4%. The combined cycle efficiency increases by 0.82–1.38%. There is a drop of 4.5–6.7% in compressor pressure ratio while the TIT increases by 0.9–7.8%. Although the steam turbine performance depends on the turbine exhaust temperature (TET) and the effectiveness of the HRSG, the effect of ambient condition is primarily manifested in the GT performance.

Power Augmentation. To boost power production and efficiency, two power augmentation methods are considered: (a) gas turbine fog inlet cooling and (b) steam injected gas turbine (STIG).

Cases 17 - 20: Combined Cycles with Inlet Fog Cooling at Four Different Ambient Conditions, Producer Gas 1. GT fog inlet cooling has been considered as an economic and effective means to augment GT power output on hot or dry days [7]. Mist or fog inlet can increase gas turbine output by cooling down the inlet air by evaporation of the fine water particles and hence reduces the compressor work. The added water increases the total mass flow rate, so it also slightly increases the output power. With increased fuel mass flow rate for using LCV producer gases, it is important to examine how the compressor is going to perform to allow more mass flow rate passing through the combustor and the turbine. The previous four cases (13–16), under different weather and humidity conditions, are simulated with GT inlet fog cooling as Cases 17–20.

Figure 3 shows a representative row of foggers in front of the compressor in the diagram. The relative humidity for each case of 17–20 is raised to 100% by assuming all the fine mist droplets are completely evaporated before entering the compressor. The water droplets are injected at 5 to 10 μm in diameter. The projected air temperature after fog cooling, but before entering the compressor, is assumed to reaching the wet bulb temperature. For example, for Case 17 after saturation fogging, the compressor inlet temperature is assumed to have reached the wet bulb temperature at 57.9°F (287.5 K).

Table 8 shows the comparison of the inlet fog cooling results with the ISO condition of Case 10. The results show inlet fog cooling will slightly increase the plant efficiency (1% to 2%) but not increase in the total output power when compared with the ISO condition. This is expected because the ISO condition is at relatively low ambient temperature, and at best, fog cooling can only reach the wet bulb temperature. However, the main purpose of employing the fog inlet cooling is not for improving the ISO condition; rather, it is for augmenting power for the hot or dry ambient conditions. Therefore, the interest is focused on comparing the cases between 13 and 16 with 17 and 20 as shown in Table 9.

Table 9 shows that inlet fog cooling consistently provides power augmentation (0.4–3.7%) at dry or hot environment; however, the efficiency of each case slightly decreases (–0.11% to –0.97%). The power augmentation is more pronounced in the dry environment (30% RH for Cases 17 and 19) than in a hot environment (Case 20). Of course, the dry and hot environment harvests the most power augmentation out of fog cooling (Case 19) but with the worst degrading of efficiency. The result of the increased power output that accompanies the increased heat rate indicates more fuel is needed to heat up the saturated air for generating additional power at a less efficient way than the dry air.

Overall, power augmentation for a small gas turbine (5 MWe) is not as good as a larger gas turbine. The use of LCV fuels further impedes the performance of fog cooling because the compressor is already heavily loaded by the increased backpressure due to excessive fuel mass flow rate. An additional increase of mass flow rate from the water vapor in a combined cycle does not augment as much power (in percentage) as the natural gas fired GT system.

Cases 21–24 (STIG): Combined Cycles with Combustor Steam Injection at Various Ambient Conditions, Producer Gas 1. Steam injected into a combustor has been regularly employed in power industry to reduce NOx emissions and for augmenting output power as well. Steam injection was used in the 1940s to boost the power output of military airplane engines [8]. In the 1970s, steam injection was employed to lower the combustion flame temperature and reduce NOx formation [9]. The performance of a steam injection cycle also depends on how the steam is produced through the heat recovery steam generator (HRSG). Four cases are simulated in STIG study with three considerations: varying ambient temperatures, employing maximum potential of steam injection, and using steam mass flow rate the same as the fogger water

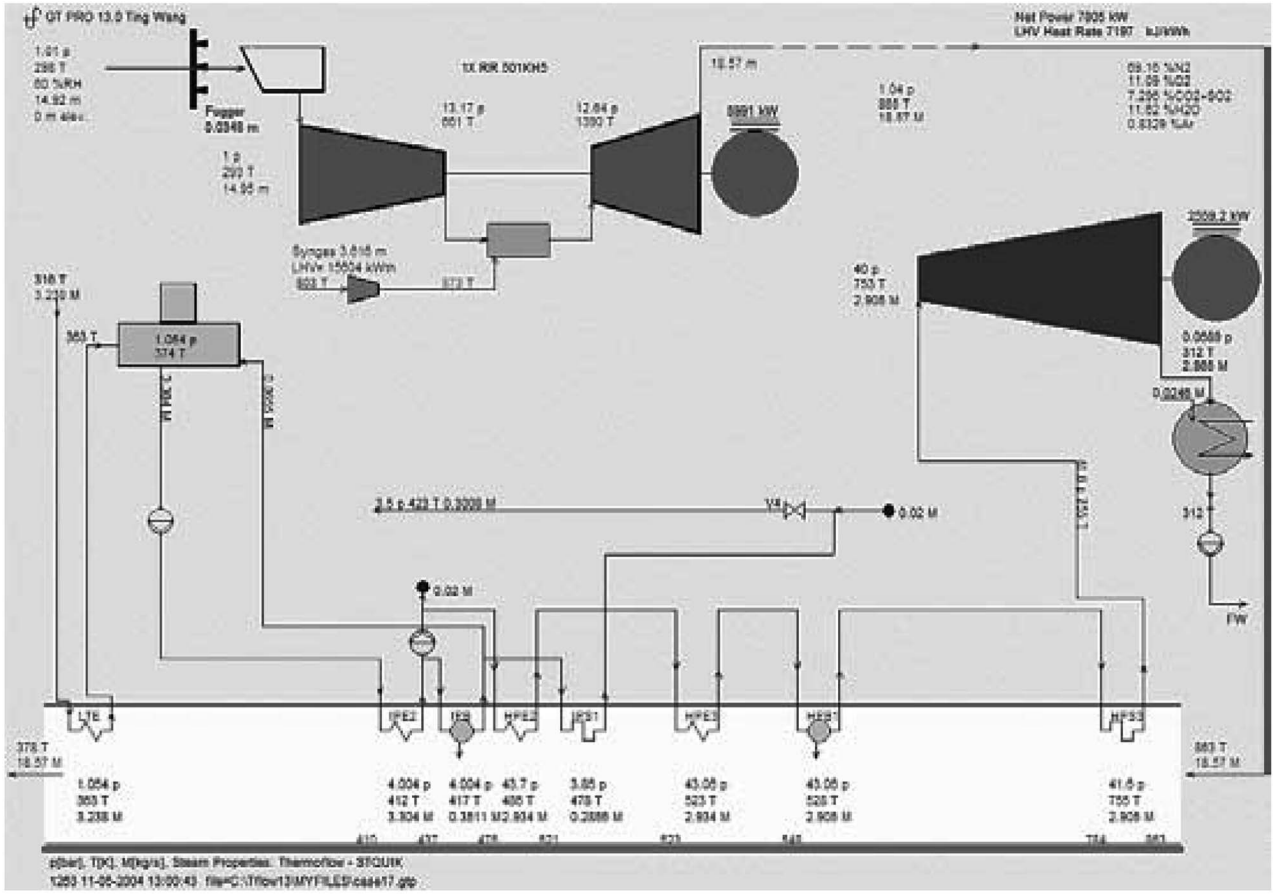


Fig. 3 THERMOFLOW schematic output for Case 17 with inlet fog cooling using producer gas 1

Table 8 Combined cycle inlet cooling cases 17–20 using Producer Gas 1

Combined Cycle	ISO	Inlet fog cooling			
Case number	10[f]	17	18	19	20
Ambient Temp (K)	288.1	298.1	298.1	305.3	305.3
(°F)	(59)	(77)	(77)	(90)	(90)
Ambient RH	0.6	0.3	0.9	0.3	0.9
GT inlet temp after fog cooling (K)	288.1	287.5	295.9	293.1	303.1
(°F)	(59)	(57.92)	(73)	(68)	(86)
RH after fog cooling	0.6	1.0	1.0	1.0	1.0
Pressure ratio	13.4	13.1	12.9	13.1	12.5
TIT (K)	1385	1390	1395	1390	1402
Air (kg/s)	15.3	15.0	14.6	14.9	14.1
Fuel LHV (MJ/kg)	4.32	4.32	4.32	4.32	4.32
Fuel HHV (MJ/kg)	5.09	5.09	5.09	5.09	5.09
Fuel (kg/s)	3.659	3.616	3.569	3.615	3.496
Fuel compressor power (MWe)	0.608	0.601	0.593	0.600	0.581
GT power (MWe)	6.067	5.991	5.911	5.990	5.783
ST power (MWe)	2.555	2.559	2.557	2.559	2.554
Net total power (MWe)	7.870	7.805	7.731	7.803	7.614
LHV GT heat rate (kJ/kW h)	9371	9377	9380	9377	9390
Total LHV heat rate (kJ/kW h)	7224	7197	7171	7197	7133
LHV GT efficiency (%)	38.42	38.41	38.38	38.39	38.34
Total LHV thermal efficiency (%)	49.84	50.02	50.20	50.02	50.47
GT power increase (%)		−1.25	−2.57	−1.27	−4.68
Total power increase (%)		−0.83	−1.77	−0.85	−3.25
GT efficiency increase (%)		−0.08	−0.10	−0.08	−0.21
Total LHV efficiency increase (%)		0.36	0.72	0.36	1.26
Total LHV heat rate increase (%)		−0.37	−0.73	−0.37	−1.26

[f]Case 10 is the reference case for comparison with other cases. See other notes in Table 6.

Table 9 Power augmentation by employing inlet fogging on cases 13–16. The results are summarized as cases 17–20.

Combined cycle	Effect of fog cooling at various ambient conditions			
Case Number	13→17	14→18	15→19	16→20
Temperature (K) and humidity (%)	298, 30%→287, 100%	298, 90%→295, 100%	305,30%→293,100%	305, 90%→303,100%
GT power increase (%)	2.25	0.51	4.76	0.59
Total power increase (%)	1.84	0.39	3.71	0.49
GT efficiency increase (%)	−0.07	0.00	0.00	0.03
Total efficiency increase (%)	−0.46	−0.12	−0.97	−0.12
Total heat rate increase (%)	0.45	0.11	0.98	0.11

mass flow rate.

Cases 21 and 22 are at medium temperature (77°F) and 90% humidity. Cases 23 and 24 are at high (90°F) temperature and 60% humidity. Cases 21 and 23 are at a lower TIT where the firing temperature is kept low while maximizing the steam recovery and hence increasing the amount of steam being injected. The lowered TIT is roughly at 84% to 85% of the designed maximum TIT of 1450 K (2150°F/946°C). Cases 22 and 24 are simulated at a higher TIT with less steam recovered, and the steam injected mass flow rate is matched with similar mass flow rate of the fogger. Under the same ambient conditions, Cases 21 and 22 are compared to Case 14 and 18, respectively, while Cases 23 and 24 are compared to Cases 16 and 19, respectively.

Figure 4 shows the GT schematic with the HRSG, which supplies steam for injection into the combustion chamber via the high pressure substream. The similar ambient condition in Case 14 (77°F, 90% RH) is now simulated with steam injection (Cases 21 and 22) into the combustion chamber to augment power under the condition of using LCV fuels. Case 21 investigates the cycle performance by maximizing the steam injection mass flow via the same HRSG. In Case 22, the amount of steam injection matches the same amount of water used for fog cooling in Case 18, which compares the effect of power augmentation between steam injection and fog inlet cooling. Cases 23 and 24 repeat the same simulation method as Case 21 (maximum steam injection mass flow) and Case 22 (matching water mass flow rate in Case 19) with the similar ambient condition of Case 16 (90°F, 30% RH).

Table 10 summarizes the performance of steam injection Cases 21–24 and their comparisons with ISO Case 10. Both Cases 21 and 23 inject the maximum steam mass flow that can be generated by the HRSG at the pressure comparable to (or a bit higher than) the combustor pressure. When the steam is diverted into the combustor, the steam turbine operates at inefficient partial load. The results of Cases 21 and 23 show that the GT power increases 4% to 5%, and GT efficiency increases 15%. However, the steam turbine output drops 91%, and the total plant performance suffers

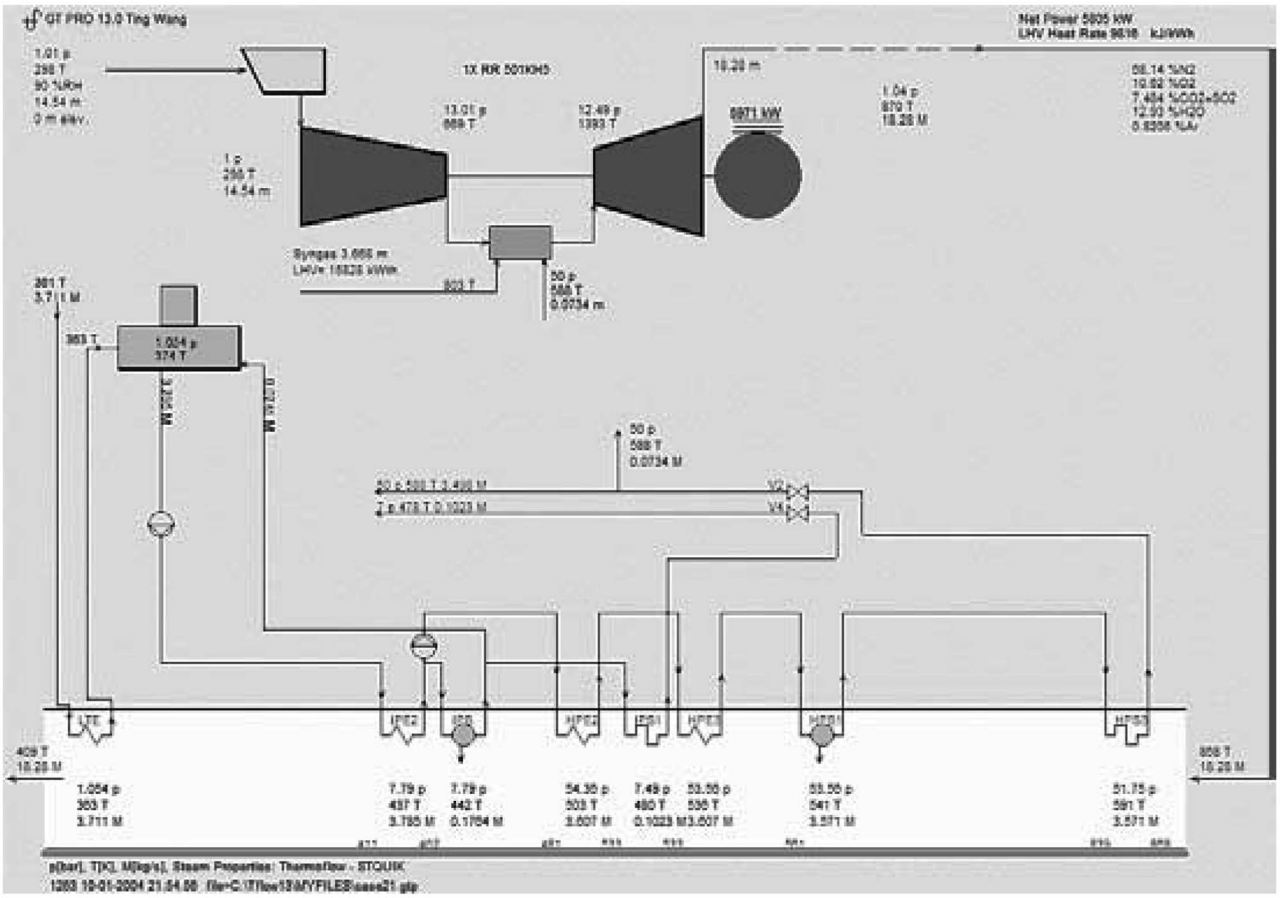


Fig. 4 THERMOFLOW schematic output for combustor steam injection in Case 21

Table 10 Summary of combined cycle steam injection performance for Cases 21–24[g]

Case Number	10[h]	21	22	23	24
Ambient Temp (K°(F))	288.1 (59)	298.1 (77)	298.1 (77)	305.3 (90)	305.3 (90)
Rel. hum.	0.6	0.9	0.9	0.3	0.3
Pressure ratio	13.4	14.3	12.8	14.0	12.6
TIT (K)	1385	1193	1396	1211	1404
Air (kg/s)	15.3	14.5	14.5	14.2	14.2
Steam injection (kg/s)	0	3.03	0.01	2.92	0.08
Fuel LHV (MJ/kg)	4.32	4.32	4.32	4.32	4.32
Fuel HHV (MJ/kg)	5.09	5.92	5.92	5.92	5.92
Fuel (kg/s)	3.659	3.324	3.553	3.302	3.470
Fuel compressor power (MWe)	0.608	0.552	0.590	0.548	0.576
GT power (MWe)	6.067	6.370	5.886	6.314	5.774
ST power (MWe)	2.555	0.236	2.565	0.230	2.484
Net total power (MWe)	7.870	5.896	7.717	5.873	7.583
LHV GT heat rate (kJ/kW h)	9371	8108	9376	8125	9337
Total LHV heat rate (kJ/kW h)	7224	9321	7152	9077	7152
GT LHV efficiency (%)	38.42	44.40	38.40	44.31	38.56
Total LHV thermal efficiency (%)	49.84	38.62	50.34	39.66	50.34
GT power increase (%)		4.99	−2.98	4.07	−4.83
Total power increase (%)		−25.08	1.94	−25.37	−3.65
GT efficiency increase (%)		15.56	−0.05	15.33	0.36
Total LHV efficiency increase (%)		−22.51	1.00	−20.43	1.00
Total LHV heat rate increase (%)		29.03	−1.00	25.65	−1.00

[g]See other notes in Table 6.
[h]Case 10 is the ISO referene case for comparisons with no steam injection.

a 25% reduction in net output power and 20–22% loss in efficiency. Basically, the increased performance in GT is not sufficient enough to make up for the poor performance in the steam turbine, and subsequently, a loss to the entire plant occurs.

Steam injection does not help Cases 22 and 24 enough to augment power in comparison with the ISO condition; however, the thermal efficiency for Cases 22 and 24 is raised by 1.0%. The slightly increased thermal efficiency of Cases 22 and 24 may be due to the use of the HRSG to recover waste heat to produce superheated steam.

Similar to the motivation for using fog inlet cooling, the purpose of using steam injection is not to augment power at the ISO condition; rather, it is for power augmentation under non-ISO conditions. The performance of Cases 21 and 23 are equally not as good when compared with Cases 14 and 16, respectively (Table 11). This indicates that maximizing steam injection is not a good practice for a combined cycle designed with optimized load shares between the duties of GT and ST.

Comparison of the inlet fog cooling and steam injection using the same amount of water mass flow (Cases 22 versus 18 and 24 versus 19) indicates that steam injection provides a minor edge over fog inlet cooling in augmenting power under both dry and humid ambient conditions. Fog inlet cooling, however, shows a better efficiency than steam injection (Cases 19 versus 24) under the dry ambient condition (30% RH). When relative humidity is high, fog cooling understandably underperforms the steam injection cases. In summary, steam performs better when expanded in the steam turbine than being injected in to the gas turbine combustor for a combined cycle system. This is especially valid for the system burning LCV fuels because the compressor is already burdened with higher backpressure. Brun, et al. [10] specifically discussed a simplified method to evaluate the principal factors that affect the aerodynamic stability of a single shaft gas turbine's axial compressor.

Table 11 Comparison of combined cycle steam injection performance for Cases 21–24 versus Cases 14, 18, 16, and 19

Combined Cycle	Steam injection versus fog cooling			
Case number	21 vs 14	22 vs 18	23 vs 16	24 vs 19
Ambient Temp (K)	298.1	298.1	305.3	305.3
Rel. hum.	0.9	0.9	0.3	0.3
GT power increase (%)	8.31	−0.42	10.42	−3.61
Total power increase (%)	−23.44	−0.18	−21.94	−2.82
GT efficiency increase (%)	15.69	0.05	15.60	0.44
Total efficiency increase (%)	−23.16	0.28	−21.48	0.64
Total heat rate increase (%)	30.13	−0.26	27.36	−0.63

Their analysis showed that when inlet and interstage water injection is combined with other factors such as LCV fuels and combustor steam injection, gas turbine compressor aerodynamic stability problems such as rotating stall and flutter will likely occur. These aerodynamic instabilities can be directly linked to blade high-cycle fatigue and possible catastrophic gas turbine failure. Furthermore, any water injection into a gas turbine (inlet, interstage, or combustor) will reduce the hot-section turbine parts life. Therefore, care must be taken to employ inlet fog cooling or combustor steam injection when LCV fuels are burned. A companion paper [11] specifically discusses the LCV fuels fired GT performance under inlet fog overspray.

Limits and Uncertainty of the THERMOFLOW Code. The uncertainty of the THERMOFLOW code consists of three major contributors:

(a) Compressor performance—When inlet fogging is employed, the potential effect of water droplets on the compressor efficiency is not simulated. The compressor's performance is not calculated from the OEM's compressor performance map, which is usually not accessible, but estimated from a compressor performance map generated by THERMOFLOW.
(b) Combustor performance—The combustor is assumed functional when burning producer gases. The combustor is treated as a black box, and the actual combustion mechanisms are not modeled. The pressure drop calculation is reasonable in the combustor.
(c) Turbine performance—To accommodate the significant increase of fuel flow, the turbine nozzle area is adjusted in the software to allow more gas to pass through. In the real engine, the turbine nozzle area cannot be adjusted by the operator; the OEM must redesign the turbine section if the same turbine efficiency is to be maintained. This could be

done by enlarging the passage tip/hub ratio or redesigning the airfoils with a higher loading factor and low solidity (i.e., reducing the turbine blade number). If the airfoils are not to be changed, the turbine nozzle area can be opened up by altering the turbine blade stagger angle. By doing so, the turbine efficiency usually drops due to altered incidence angle and lift coefficient. In the THERMOFLOW simulation, the turbine is assumed to have the same isentropic efficiency when the turbine nozzle area is opened to accommodate more through-flow.

The heat balance errors of all the cases are less than 0.05%. For NG cases, the uncertainties of plant net power and thermal efficiency are estimated to be within 3%. For cases fired with producers' gases, the uncertainty is estimated within 5%, assuming the assumptions made in the software hold valid and the software limits are understood.

Conclusions

This study conducts performance analyses of small power plants fueled with LCV producer gases derived from biomass gasification process. The analyses are performed using the proven commercial code THERMOFLOW (release 13).

Three biomass derived producer gases with low heating values (LHV) of 4.32 MJ/kg, 7.4 MJ/kg, and 10.31 MJ/kg respectively, are used. To achieve the rated power, the LCV fuels must be supplied at a flow rate approximately 4.6–10 times more than a natural gas fired gas turbine. These increased mass flow rates resulted in higher backpressures and increased loads on the compressor blades. To accommodate the backpressure problem, the first-stage turbine nozzle openings are widened to achieve the design compression ratio of the gas turbine engine. When compared with the natural gas fueled system, the results show that typically the producer gas fueled system result in higher net output power and increased plant efficiency for 5 MWe plants based on fuel supplied from gasifier at 10.34 bars 803 K. The best performance occurs when the designed pressure ratio with widened nozzle openings is achieved, even though the TIT are relatively low (Cases 2c and 2d).

When the heating value increases the GT net output power and efficiency are all higher than the natural gas fueled system in a subsequently reduced amount.

The ambient temperature and relative humidity plays a significant role in the performance of a gas turbine system. The result of 5 MWe inlet fog cooling increases GT output power from 4.76% to 0.5% (Cases 17, 18, 19, and 20) with negligible GT efficiency changes. Inlet fog cooling provides lower power augmentation for the combined cycle than for a simple GT cycle, ranging from 3.7% to 0.39%, and adversely affects the CC efficiency ranging from −0.1% to −1%. An additional burden from the fog cooling with the increased mass flow rate results in an undesirable compressor performance and reduced TIT, both factors contribute to a marginal GT power augmentation and an adverse impact to total plant efficiency when LCV fuels are used.

Comparison between inlet fog cooling and steam injection using the same amount of water mass flow indicates that steam injection is doing worse than inlet fog cooling in augmenting power output when LCV fuels are used. The results show that the steam performs better when the steam is to be expanded in the steam turbine than being injected into the gas turbine combustor for a combined cycle system. This is especially valid for the system burned with LCV fuels because the compressor is already burdened with increased backpressure. Maximizing steam injection, at the expense of supplying the steam to the steam turbine, significantly reduces both the efficiency and the output power of the combined cycle.

The above-noted conclusions are based on the fuel compressor power being calculated on the inlet condition at 803 K and 10.34 bars. If the fuel compressor power is calculated based on 298 K and 1 atm, and is entirely treated as a parasitic power, the net power output and thermal efficiency will need to be re-evaluated.

This study indicates that the performance of GT and CC systems fueled by the LCV fuels could be very different from the familiar behavior of natural gas fired systems. Care must be taken if on-shelf GTs are used to burn LCV fuels.

Acknowledgment

This study is supported by the Louisiana Governor's Energy Initiative via the Clean Power and Energy Research Consortium (CPERC) and administered by the Louisiana Board of Regents. The comments from THERMOFLOW staff are appreciated.

Nomenclature

BIGCC = biomass integrated gasification combined cycle
CC = combined cycle
GT = gas turbine
HHV = high heating value (MJ/kg, Btu/lbm)
HP = high pressure
HRSG = heat recovery steam generator
IGV = inlet guide vane
IP = intermediate pressure
LCV = low calorific value
LHV = low heating value (MJ/kg, Btu/lbm)
LP = low pressure
m = mass flow rate (kg/s)
MWe = megawatts electricity
NG = natural gas
NOx = nitrous oxides
η = thermal efficiency
p = pressure (bar, Pa, psi)
PG = producer gas
RAM = reliability, availability, and maintainability
RH = relative humidity (%)
ST = steam turbine
T = temperature (K, °C, °F)
TET = turbine exit temperature (K, °C, °F)
TIT = turbine inlet temperature (K, °C, °F)

References

[1] Bedi, E., 2001, "Biomass Energy," http://www.geocities.com/dieret/re/Biomass/biomass.html
[2] Swanikemp, R., 2001, "Project Developers Consider New Solid Fuels, New Technologies," Power, **145**(2), pp. 35–42.
[3] Layne, A. W., 2001, "Advanced Turbine Systems Program," National Energy Technology Laboratory, http://www.netl.doe.gov/publications/factsheets/program/prog002.pdf
[4] Yap, M. R., and Wang, T., 2004, "Simulation of Producer Gas Fueled Power Plants," ECCC Report No. 2004-07, Energy Conversion and Conservation Center, University of New Orleans.
[5] THERMOFLOW Inc., 2004, THERMOFLOW manual, release 13.
[6] Ishimura, D. M., Kinoshita, C. M., Masutani, S. M., and Turn, S. Q., 1999, "Cycle Analyses of 5 and 20 MWe Biomass Gasifier-Based Electric Power Stations in Hawaii," ASME J. Eng. Gas Turbines Power, **121**, pp. 25–30.
[7] Nicholson, A., 2004, "Cooling Potential Across The US," MEE Industries Inc., http://www.meefog.com/turbine/gtps-cp.html
[8] Cohen, C., Rogers, G. F. C., and Saravanamauttoo, H. I. H., 1999, *Gas Turbine Theory*, 4th ed., Longman Group Limited, Essex, UK.
[9] Petrotech, July 1999, "Gas Turbine Steam Injection System for NO_x Reduction," ⟨www.petrotechinc.com/pdf/94023.pdf⟩, Product Bulletin No. 94023.
[10] Brun, K., Kurz, R., and Simmons, H. R., 2005, "Aerodynamic Instability and Life Limiting Effects of Inlet And Interstage Water Injection Into Gas Turbines," ASME Paper No. GT2005-68007.
[11] Khan, J., and Wang, T., 2006, "Fog and Overspray Cooling for Gas Turbine Systems Using Low Calorific Value Fuels," ASME Paper No. GT2006-90396.

Dynamics of a High-Frequency Fuel Actuator and its Applications for Combustion Instability Control

Tongxun Yi

Ephraim J. Gutmark

Department of Aerospace Engineering and Engineering Mechanics, University of Cincinnati, Cincinnati, OH 45220-0070

The present paper performs complementary experimental and theoretical investigations of a pump-style, high-frequency, magnetostrictive fuel actuator, which achieves fuel modulations by periodically "pushing" fuel out of a piston-cylinder unit instead of by modulating the flow area. The low-order models are developed to identify relevant parameters and investigate their influences on fuel modulations. This fuel actuator is proprietary, its detailed internal structure and dimension are not available, so only qualitative comparisons between the model predictions and experiments are made. Experiments validate the trend of the model predictions. A system-identification-based LQG controller is designed to quickly suppress strong interferences of fuel modulations with the mean fuel flow rate. Improvements to the fuel setup have been made according to the model predictions, which have been experimentally shown to be beneficial to combustion instability control. [DOI: 10.1115/1.2718558]

1 Introduction

Dry-low-emission (DLE) combustion generates very low NO_x by premixing fuel with a very large percentage of combustion air [1]. However, DLE combustion is susceptible to combustion instability, which is probably caused by equivalence ratio variations and insufficient acoustic damping. Both passive and active control strategies can be implemented to mitigate the detrimental, large-amplitude combustion oscillations. Typical passive control strategies include the quarter-wave tubes [2], acoustic liners [3], modifying the combustor geometry, changing the fuel injection locations [4], and modifying the fuel distribution pattern. Compared with the passive control strategies, the active ones are usually capable of working in broader ranges, and may only require modifications external to an engine. However, the shortage of a reliable, high frequency, proportional fuel actuator has become one of the major technical challenges to combustion instability control. Several high frequency fuel actuators have been reported in the literature [5–7], all of which achieve fuel modulations by periodically modulating the flow area.

The present paper investigates a pump-style, high frequency, magnetostrictive fuel actuator, which achieves fuel modulations by periodically pumping fuel out of a piston-cylinder unit (PCU). Complementary experimental and theoretical investigations have been performed to improve the fuel setup to achieve large fuel modulations. A LQG pulse-width-modulation controller is designed to quickly reject the interferences of fuel modulations with the mean fuel flow rate. With the improved fuel setup and the LQG mean flow controller, successful control of combustion instability is demonstrated on an unstable gas turbine combustion simulator.

2 Experimental Setup

2.1 Combustion Rig. An ideal combustion simulator should have similar reacting flow patterns as those in a generic gas turbine combustor. Several similarity criteria were discussed by Penner in the 1950s [8]. Figure 1 shows the atmospheric combustion rig featured with the GE/Goodrich/UC Triple-Annular-Research-Swirler (TARS). The rig is vertically set up. Heated combustion air enters the combustion rig from the bottom. The air flow rate is measured with a thermal wire digital flow meter with a measurement uncertainty of 0.1%. A 0.66-m-long quartz tube with an internal diameter of 0.15 m is used as the combustion chamber. A water-cooled Kistler dynamic pressure sensor (7061B) is installed at 0.08 m above the dump plane, where pressure oscillations are most intense. The TARS has three air swirlers, with the inner swirl angle 30 deg (clockwise), the middle one 45 deg (clockwise), and the outer one 55 deg (counterclockwise). The TARS has two fuel lines, namely, the main fuel and the pilot fuel. The combustor is fueled with turpentine. The airside pressure drop across the TARS is about 3–9% of the plenum pressure

2.2 The High-Frequency Fuel Actuator. Figure 2 shows the piston-cylinder unit (PCU) of the high-frequency fuel actuator. The fuel supply is a fuel accumulator with a capacity of 0.012 m^3, which is about 7600 times larger than typical fuel modulations within one forcing cycle, so the fuel pressure within the accumulator is assumed to be constant. The mean fuel flow rate is varied when the piston slides across the exit orifice of the cylinder. The piston movement is controlled by a stepper motor with a pulse-width-modulation scheme. The mean fuel flow rate is computed from the pressure drop across an orifice installed downstream of the fuel injector. The pressure drop is measured with a Sensotec differential pressure transducer. A lowpass filter with a cutoff frequency of 1 Hz is used to obtain the mean fuel flow rate. Fuel modulations are achieved by "pushing" fuel out of the PCU using a Terfenol-D rod, which extends or contracts with the alternating magnetic fields generated by an alternating current. The alternating current is generated by a Titan oscillator, which amplifies the fuel modulation signal. One can infer from Fig. 2 that, the mechanisms for fuel modulations are mechanically coupled with those for mean flow control, which is undesirable.

2.3 Instrumentation. The mean fuel flow rate is controlled with a Dspace Controldesk with a Dspace board CP1104. Data acquisition is done using Labview 7.0 and a NI-DAQ board PCI-MIO-16XE-10. The sampling frequency is 5000 Hz, and the data length is 10 s. An F150 Precision Thermometer from *ASL* and type B thermocouples are used for temperature measurement.

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the JOURNAL OF TURBOMACHINERY. Manuscript received December 29, 2004; final manuscript received December 20, 2006. Review conducted by William Rhoden.

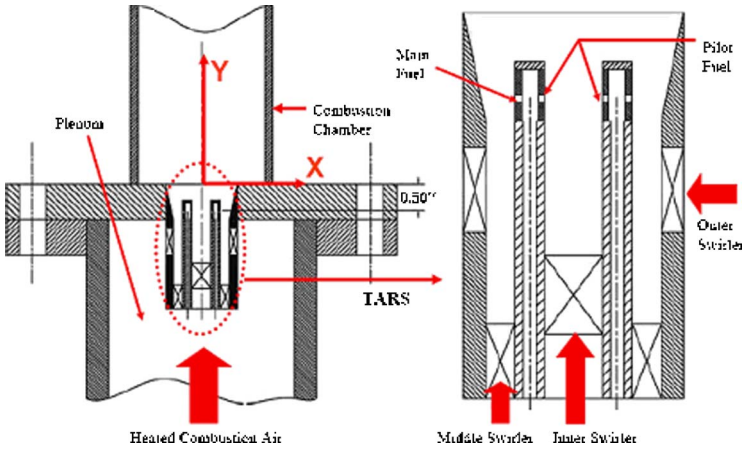


Fig. 1 Atmospheric combustion rig with the TARS

3 Low-Order Modeling

Low-order models are developed to identify the relevant parameters and investigate their influences on fuel modulations. This fuel actuator is proprietary, its detailed internal structure and dimension are not available, so the models are more suitable for qualitative predictions rather than quantitative predictions. Figure 3 shows a sketch of the simplified fluid circuit, which has included the main components of the fuel setup.

3.1 Governing Equations. Flow within the fuel tubing is modeled as 1D inviscid flow based on the following assumptions:

- Axisymmetric flow;
- Short tubing length compared with the acoustic wavelength;
- Negligibly small viscous stress;
- Low-Mach number flow.

Note that pressure wave initiated by the piston movement propagates along the tubing with a finite sound speed, $a=\sqrt{K/\rho}$, where K is the isentropic modulus of fuel. The assumption of short tubing allows the velocity and acceleration within the tubing to be treated as spatially uniform, which considerably simplifies modeling. Applying unsteady Bernoulli's equations to flow within the upstream and downstream tubing, respectively [9], one gets

$$\rho L_1 \frac{du_1}{dt} + p_c = p_0 \tag{1}$$

$$\rho L_2 \frac{du_2}{dt} + p_v = p_m \tag{2}$$

Note that p_v and p_m denote the fuel pressure immediately downstream of the PCU and immediately upstream of the exit orifice, respectively. The exit orifice of the PCU, which consists of a circle of small holes, is modeled as a resistive element,

$$\dot{Q} = A_e(x_m - h_0)\sqrt{2(p_c - p_v)/\rho} = A_{t2}u_2 \tag{3}$$

Finite compressibility of fuel is considered for fuel within the PCU,

$$\frac{d(\rho A_c x_m)}{dt} = \rho(A_{t1}u_1 - \dot{Q}) \rightarrow A_c x_m \frac{d\rho}{dt} = \rho\left(A_{t1}u_1 - \dot{Q} - A_c\frac{dx_m}{dt}\right) \tag{4}$$

The exit orifice at the end of the downstream tubing is modeled as a resistive element (more discussions in Sec. 4.1),

$$\dot{Q} = A_o\sqrt{2(p_m - p_a)/\rho} = A_{t2}u_2 \tag{5}$$

3.2 Steady State Solutions. At steady states, the time derivatives in Eqs. (1)–(3) vanish, resulting in the following relationship:

$$p_{c0} = p_0 \quad p_{v0} = p_{m0} = (A_{v0}^2 P_0 + A_o^2 P_{\text{atm}})/(A_{v0}^2 + A_o^2) \tag{6}$$

$$\dot{Q}_0 = A_{\text{eq}}\sqrt{2(P_0 - P_{\text{atm}})/\rho} \tag{7}$$

For the convenience of expression, we denote $A_v = A_e(x_m - h_0)$, $A_{v0} = A_e(x_{m0} - h_0)$, and $A_{\text{eq}} = A_{v0}A_o/\sqrt{A_{v0}^2 + A_o^2}$. x_m and h_0 are shown in Fig. 3.

3.3 Linearization [10]. Linearizing Eqs. (1)–(3) and Eq. (5), one gets

$$\rho L_1 \frac{du_1'}{dt} + p_c' = 0 \tag{8}$$

$$\rho L_2 \frac{du_2'}{dt} = p_v' - p_m' \tag{9}$$

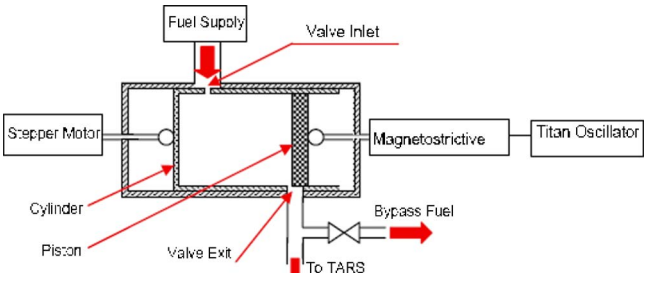


Fig. 2 Structure of the PCU

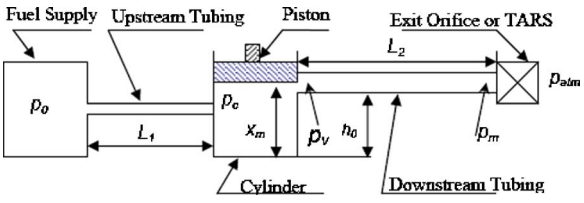


Fig. 3 Sketch of the simplified fluidic circuit

$$\frac{\dot{Q}'}{\dot{Q}_0} = \frac{p_c' - p_v'}{2(p_{c0} - p_{v0})} = \frac{p_m'}{2(p_{m0} - p_{\text{atm}})} \tag{10}$$

In Eq. (10), the effects of the PCU exit flow area on fuel modulations are neglected. For fuel modulations above 50 Hz, it is the pressure dynamics within the PCU that determines fuel modulations. With the piston quickly moving downward, despite a decreasing exit flow area of the PCU, the fuel flow rate out of the PCU still increases because of the pressure rise within the PCU. However, if the piston quickly moves downward and then stops, the fuel flow rate will finally reach a lower value than before. This suggests that the fuel actuator is of nonminimum phase [10]. "Nonminimum-phase systems are slow in response because of their faulty behavior at the start of response. In most practical control systems, excessive phase lag should be carefully avoided. In designing a system, if fast speed of response is of primary importance, we should not use nonminimum-phase components [10]." To avoid the nonminimum phase, the designer can simply move the mechanisms for mean flow control out of the PCU.

Equation (4) is linearized as,

$$A_c x_{m0} \frac{d\rho'}{dt} = \rho\left(A_{t1} u_1' - \dot{Q}' - A_c \frac{dx_m'}{dt}\right) \tag{11}$$

Note that the isentropic modulus of fuel is defined as [11]

$$\frac{d\rho'}{dt} = \frac{\rho}{K}\frac{dp_c'}{dt} \tag{12}$$

By combining Eqs. (11) and (12), pressure dynamics within the PCU can be written as

$$\frac{A_c x_{m0}}{K}\frac{dp_c'}{dt} = \left(A_{t1} u_1' - \dot{Q}' - A_c \frac{dx_m'}{dt}\right) \tag{13}$$

From Eqs. (9) and (10), one can derive the transfer function relating the volumetric fuel flow rate to the pressure dynamics within the PCU,

$$\frac{\dot{Q}(s)}{P_c(s)} = \frac{1}{\rho L_2 s/A_{t2} + 2(P_0 - P_{\text{atm}})/\dot{Q}_0} = \frac{1}{\rho L_2 s/A_{t2} + \rho\dot{Q}/A_{\text{eq}}^2} \tag{14}$$

From Eqs. (8), (13), and (14), one can derive the transfer function relating the fuel modulations to the piston movement,

$$\frac{\dot{M}(s)}{X_m(s)} = \frac{-s}{\dfrac{x_{m0}L_2 s^2}{A_{t2}K}\left(1 + \dfrac{\dot{Q}_0 A_{t2}}{A_{\text{eq}}^2 L_2 s}\right) + \dfrac{1}{\rho A_c}\left(\dfrac{L_2}{L_1}\dfrac{A_{t1}}{A_{t2}} + 1 + \dfrac{\dot{Q}_0 A_{t1}}{A_{\text{eq}}^2 L_1 s}\right)} \tag{15}$$

Equation (15) shows that fuel modulations are determined by the PCU geometry, the operating conditions, and the piston movement. The piston movement is controlled by the electromagnetic circuit whose dynamics is out of the scope of the present paper.

3.4 Simplifying Eq. (15). In most of our experiments, $A_{t1} = A_{t2} = 7.1e{-}6\ \text{m}^3$, $\dot{Q}_0 < 6.4E{-}6\ \text{m}^3/\text{s}$, $L_{1,2} > 0.33$ m, and $\omega > 314$ rad/s. Further assume $A_v = 2.5E{-}6\ \text{m}^2$, then $|\dot{Q}_0 A_{t1,2}/A_{\text{eq}}^2 L s|_{s=j\omega} < 0.07$. This allows Eq. (15) to be simplified as

$$\frac{\dot{M}(s)}{X_m(s)} = \frac{-s}{\dfrac{x_{m0}L_2}{A_{t2}K}s^2 + \dfrac{1}{\rho A_c}\left(\dfrac{L_2}{L_1}\dfrac{A_{t1}}{A_{t2}} + 1\right)} \tag{16}$$

Equation (16) describes a second-order Helmholtz resonator with the resonant frequency,

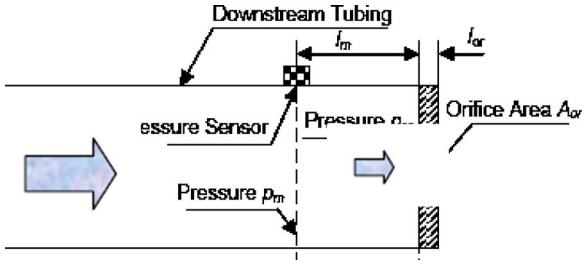


Fig. 4 Fuel flow rate measurement; $l_{or} = 1E{-}3$ m and $A_{or} = 2.5E{-}6\ \text{m}^2$

$$\omega_r = \sqrt{\frac{K}{\rho}\frac{A_{t2}}{A_c}\frac{1}{x_{m0}}\left(\frac{1}{L_1}\frac{A_{t1}}{A_{t2}} + \frac{1}{L_2}\right)} = a\sqrt{\frac{A_{t2}}{A_c}\frac{1}{x_{m0}}\left(\frac{1}{L_1}\frac{A_{t1}}{A_{t2}} + \frac{1}{L_2}\right)} \tag{17}$$

With fuel pressure around 1378–2756 kPa, $a \approx 1170$ m/s. If we assume $x_{m0} \approx 4E{-}3$ m, $A_{t2}/A_c \approx 1$, $A_{t1}/A_{t2} = 1$, and $L_1 = L_2 = 0.33$ m, the resonant frequency is computed as 7800 Hz. Note that active combustion control is usually considered for pressure pulsations below 1 kHz.

3.5 Interpreting Eq. (16). For fuel modulations below 1 kHz, the gain of the transfer function (Eq. (16)) can be expressed as

$$\left|\frac{\dot{M}(s)}{X_m(s)}\right|_{\substack{s=j\omega \\ \omega<2000\pi}} = \frac{\omega}{\dfrac{1}{\rho A_c}\left(\dfrac{L_2}{L_1}\dfrac{A_{t1}}{A_{t2}} + 1\right) - \dfrac{x_{m0}L_2}{A_{t2}K}\omega^2} > 0 \tag{18}$$

To achieve large fuel modulations, one needs to maximize $x_{m0}L_2/A_{t2}K$ and/or minimize $(1/\rho A_c)((L_2A_{t1}/L_1A_{t2}) + 1)$. This can be done as follows:

- Design a PCU with a larger cross-sectional area A_c. A larger A_c results in a smaller $(1/\rho A_c)((L_2A_{t1}/L_1A_{t2}) + 1)$. With a larger A_c, more fuel will be pushed out of the PCU for the same amplitude of the piston movement.
- Design a PCU with a larger x_{m0}. A larger x_{m0} results in a larger $x_{m0}L_2/A_{t2}K$.
- Use longer upstream tubing. A larger L_1 results in a smaller $(1/\rho A_c)((L_2A_{t1}/L_1A_{t2}) + 1)$. With a larger L_1, the fuel inertia within the upstream tubing is larger, so fuel can be more easily pushed downstream rather than upstream.
- Using upstream tubing with a smaller A_{t1}. A smaller A_{t1} results in a smaller $(1/\rho A_c)((L_2A_{t1}/L_1A_{t2}) + 1)$.
- Use shorter downstream tubing. A shorter L_2 results in a smaller $(1/\rho A_c)((L_2A_{t1}/L_1A_{t2}) + 1)$. With a shorter L_2, the fuel inertia within the downstream tubing is smaller, so fuel can be more easily pushed downstream than upstream.
- Use downstream tubing with a larger A_{t2}. A larger A_{t2} results in a smaller $(1/\rho A_c)((L_2A_{t1}/L_1A_{t2}) + 1)$.
- Use a large mean fuel flow rate out of the PCU. This is because a larger $\dot{Q}_0$ implies a larger A_{t2}. A bypass fuel passage shown in Fig. 2 can be implemented to increase the fuel flow rate out of the PCU.

Many of the above predictions have been verified by experiments, as will be shown below.

4 Experimental Results

4.1 Measurement of Fuel Modulations. Fuel modulations are determined from the pressure drop across the exit orifice, as shown in Fig. 4. The pressure drop is measured at a distance of l_m upstream of the exit orifice with a Sensotec differential pressure sensor. Since l_m is much less than the acoustic wavelength, one

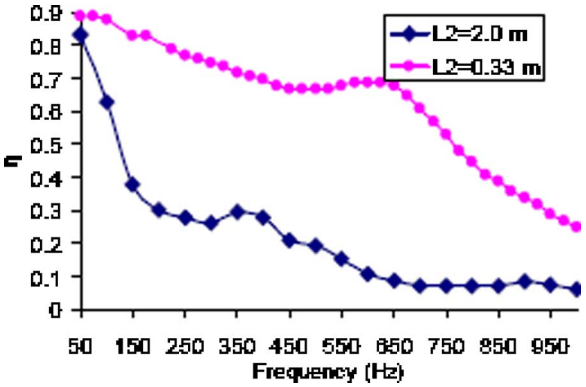


Fig. 5 Effects of L_2 on fuel modulations

can assume $p_m \approx p_{or}$. Here p_{or} denotes the fuel pressure immediately upstream of the exit orifice. Applying the unsteady Bernoulli's equations to flow across the exit orifice, one gets,

$$\rho \frac{du_o}{dt} l_{or} + \frac{1}{2}(1 + c_f)\rho u_o^2 + p_{atm} = p_{or} \approx p_m \tag{19}$$

The friction coefficient of the exit orifice, c_f, is experimentally determined to be 0.66 at the flow rate of 2.5 g/s. It is worthwhile to notice that, c_f may depend strongly on the flow Reynolds number, the channel geometry, and the forcing frequency. For simplicity, a constant c_f is assumed. The first term on the left-hand side of Eq. (19) corresponds to the flow acceleration pressure, which is denoted as p_{ac}; the second term corresponds to the pressure associated with the fluid motion and the flow resistance, which is denoted as p_{mp}. In Sec. 3.1, we simply mention that the exit orifice is treated as a resistive element. Here we justify this assumption. Assume the velocity within the orifice as,

$$u_o = u_{o0}[1 + \alpha \sin(\omega t + \psi)] \quad \alpha > 0 \tag{20}$$

α quantifies the fuel modulation level, which is fairly large at low frequencies and just moderate at high frequencies. u_{o0} denotes the mean velocity, and ψ denotes the initial phase. The ratio of p_{mp} to p_{ac} can be estimated as

$$|p_{mp}/p_{ac}| = \left| \frac{1}{2}(1 + c_f)\rho u_o^2 \Big/ \rho \frac{du_o}{dt} l_{or} \right| \propto 2\pi(1 + c_f)(1 + 0.5\alpha^2) u_{o0}/\sqrt{2}\alpha l_{or}\omega \tag{21}$$

Note that $|p_{mp}/p_{ac}|_{u_{o0}>10} \gg 1$ for $u_{o0} > 10$ m/s. This allows the exit orifice to be treated as a resistive element, and allows fuel modulations, $\dot{m}'$, to be determined as

$$\dot{m}' = A_{or}[\sqrt{2\rho(p_m - p_{atm})/(1 + c_f)} - \sqrt{2\rho(p_{m0} - p_{atm})/(1 + c_f)}] = \dot{m}_0[\sqrt{[p_m(t) - p_{atm}]/(p_{m0} - p_{atm})} - 1] \tag{22}$$

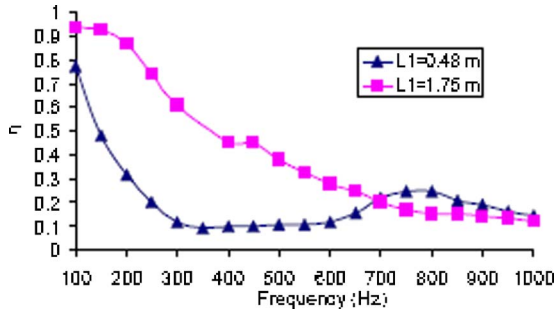


Fig. 6 Effects of L_1 on fuel modulation

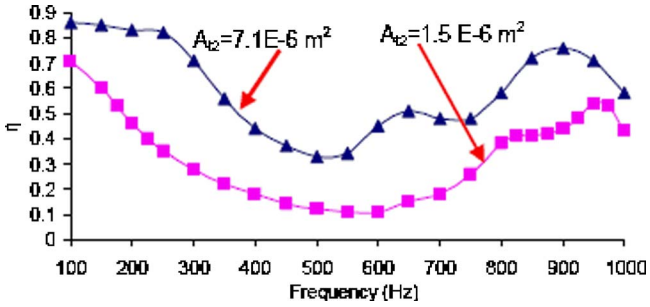


Fig. 7 Effects of A_{t2} on fuel modulations

4.2 Effects of L_2. Figure 5 shows the experimental results that compare the effects of the downstream tubing length on fuel modulations. The fuel modulation ratio, η, is defined as the ratio of the fuel modulation RMS to $\dot{m}_0$. The fuel modulation RMS is obtained by filtering $\dot{m}$ with a fourth order Butterworth bandpass filter within $[F-25, F+25]$ Hz, where F is the forcing frequency. Although shorter downstream tubing favors larger fuel modulations, it is not advisable to install the fuel actuator immediately upstream of the burner, due to its interference with the air flow into the combustor and the relatively high temperature [12]. In this experiment, $\dot{m}_0$=3.5 g/s, p_0=250 psi, L_1=0.45 m, $A_{t1}=A_{t2}=7.1E-6$ m^2, and the forcing amplitude is 0.75 V. The bypass fuel passage is used.

4.3 Effects of L_1. Figure 6 shows the experimental results that compare the effects of the upstream tubing length on fuel modulations. Longer upstream tubing favors larger fuel modulations, as predicted by the low-order models. For all results reported in this figure, $\dot{m}_0$=2.5 g/s, p_0=250 psi, L_2=0.91 m, $A_{t1}=A_{t2}=7.1E-6$ m^2, and the forcing amplitude is 0.75 V. The bypass fuel passage is used.

4.4 Effects of A_{t2}. Figure 7 shows that a small A_{t2} reduces fuel modulations, as predicted by the low-order models. Large fuel modulations are achieved around 900 Hz, which may be associated with the acoustic resonance. As a simple analysis, one may view the exit orifice as a pressure antinode, and treat the exit orifice of the PCU as a pressure node, then the resonant frequency of the three-quarter wave mode of the downstream tubing is estimated to be about 900 Hz. For all results reported in this figure, $\dot{m}_0$=3.5 g/s, p_0=250 psi, L_1=0.45 m, L_2=0.91 m, $A_{t1}=7.1E-6$ m^2, and the forcing amplitude is 0.75 V. The bypass fuel passage is used.

4.5 Effects of $\dot{m}_0$. Figure 8 shows that larger fuel modulations are achieved at $\dot{m}_0$=3.5 g/s than at $\dot{m}_0$=2.5 g/s. The fact that fuel modulations increase with $\dot{m}_0$ is quite encouraging, since the fuel flow rate required by an aero engine is typically large [6]. Note

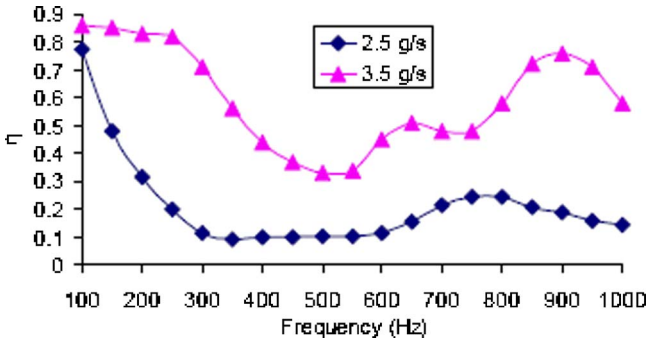


Fig. 8 Effects of $\dot{m}_0$ on fuel modulation

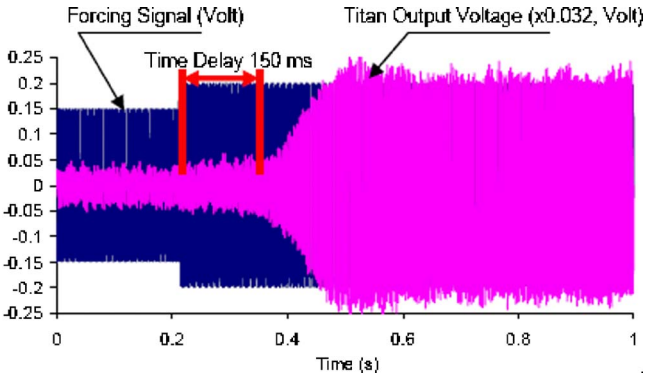


Fig. 9 The slow responses and the actuating dead zone of the Titan oscillator

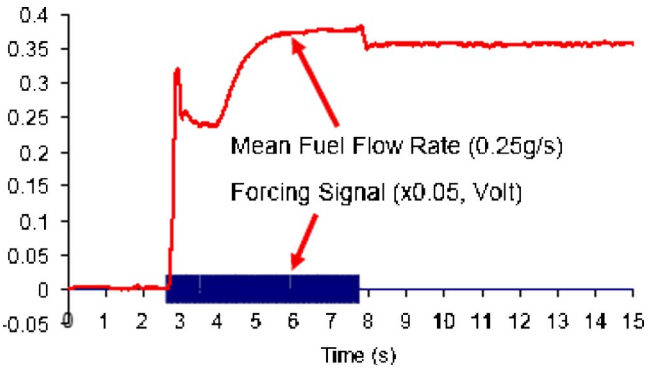


Fig. 11 Strong interferences of fuel modulations with $\dot{m}_0$

that the fuel flow rate can be significantly increased by implementing a bypass fuel passage downstream of the PCU, as shown in Fig. 2. One can see from Figs. 7 and Fig. 8, when $\dot{m}_0$ is decreased from 3.5 g/s to 2.5 g/s, the resonant frequency is shifted from 900 Hz to 750 Hz. This may be associated with the variations in the acoustic impedances at the beginning of the downstream tubing. The exit flow area of the PCU is smaller at a smaller $\dot{m}_0$. For all tests reported in Fig. 8, p_0=250 psi, L_1 =0.45 m, L_2=0.91 m, A_{t1}=A_{t2}=7.1E−6 m^2, and the forcing amplitude is 0.75 V. The bypass fuel passage is used.

4.6 Dynamics of the Titan Oscillator. The Titan oscillator is to amplify the control signal, which is usually within ±0.75 V. Dynamics of the Titan oscillator is rather slow. As shown in Fig. 9, it takes about 310 ms for the Titan output voltage to reach another fully developed oscillating state. The long transient implies a large time delay. It is also noticed that the Titan oscillator has almost no response to the forcing signal with amplitude below 0.15 V. This is referred to as the actuating "dead zone" [10]. Both the time delay and the actuating dead zone make combustion control extremely difficult, thus should be avoided as much as possible. However, we notice that, by adding white noise to the forcing signal, dynamics of the Titan oscillator can be somewhat improved. Limited by space, these results will not be shown here.

4.7 Effects of Bubbles. Air bubbles will reside within the fuel system if air is not completely purged out of the PCU at the beginning of the experiments. Fuel bubbles may also be formed within the fuel system either by flow separations or by large pressure drops within the PCU. Figure 10 shows that bubbles decrease fuel modulations more than 15 times. In addition, the bubbles also introduce phase shift to the fuel modulations. The "normalized fuel modulations" are defined as $\dot{m}'/\dot{m}_0$. For all results presented in this paper except those shown in Fig. 10, air is removed from the fuel system with a vacuum machine at the beginning of the experiments.

5 Mean Flow Regulation

5.1 Necessity of the Fast Regulation of $\dot{m}_0$. As mentioned before, the mechanisms for fuel modulations are strong coupled with those for mean flow control. Figure 11 shows that, due to the strong interferences imposed by fuel modulations, $\dot{m}_0$ is increased by 14%. Figure 12 shows that combustion instability control fails if $\dot{m}_0$ cannot be quickly regulated. In Fig. 12, an empirical proportional-derivative (PD) controller is used to regulate $\dot{m}_0$. Theoretically a PD controller is capable of very fast responses. However, the exit flow area of the PCU consists of a circle of small holes whose diameters is about ten times smaller than the moving range of the stepper motor. If the control signal is large, which is required to achieve fast responses, the stepper motor may move into the actuating "dead zone."

5.2 Closed-Loop System Identification. A system-identification-based LQG controller is designed for fast regulation of $\dot{m}_0$. Figure 13 shows the Matlab Simulink model for system identification. A PD controller with a proportional gain of 0.8 and a derivative gain of 0.1 is used to stabilize $\dot{m}_0$. The white noise has a standard deviation of 1.5, and the period of the repeating sequence is 50 ms. By following the standard procedures in Ref. [13], a transfer function is derived as follows:

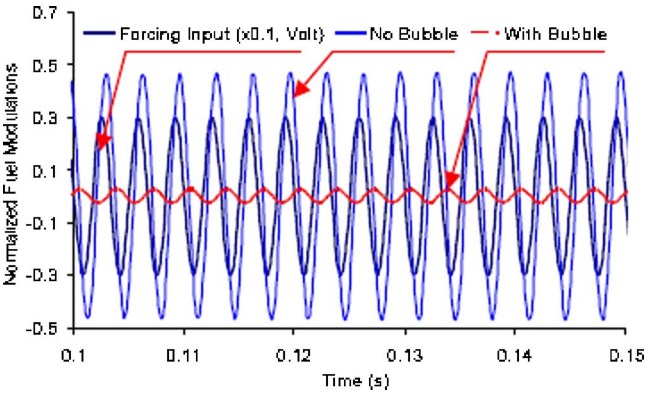


Fig. 10 The adverse effects of bubbles on fuel modulations

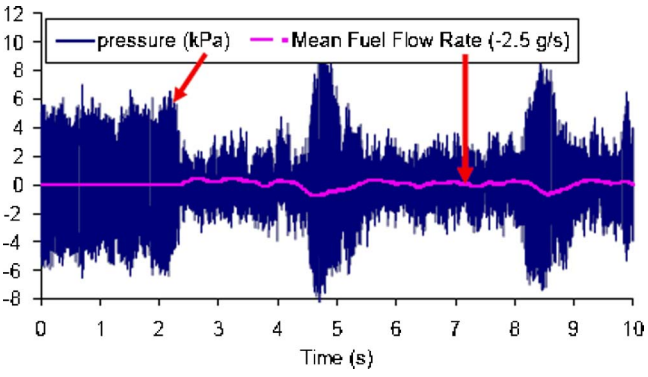


Fig. 12 Failure in combustion instability control due to slow regulation of $\dot{m}_0$

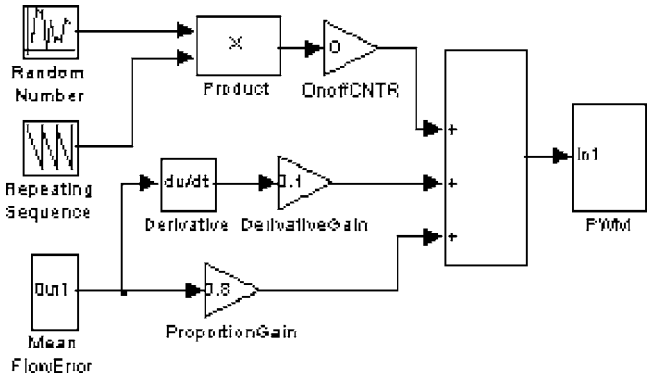


Fig. 13 Simulink model for closed-loop system identification

$$\frac{\dot{M}_0}{U(s)} = -\frac{0.004198s^3 - 1389s^2 - 6.284E6s - 6.431E10}{s^4 + 7287s^3 + 4.155E7s^2 + 2.744E11s + 3.033E9} \tag{23}$$

Obviously this model contains zeros in the right half phase plane, implying that the actuator is of nonminimum phase [10].

5.3 LQG Control Design. Based on the transfer function in Eq. (23), a LQG controller is designed as follows [10]:

$$\dot{\hat{X}} = A_c X + Gy$$

$$u = -B^T X$$

$$G = 0.001\begin{pmatrix} 0.3067 \\ 0.005 \\ -0.0027 \\ -0.0013 \end{pmatrix} \quad B = 1000\begin{pmatrix} -7.0633 \\ -3.3158 \\ 8.3638 \\ 3.4277 \end{pmatrix} \quad A_c$$

$$= 1000\begin{pmatrix} -0.0236 & -0.0171 & 0.0295 & 0.0078 \\ -1.2683 & -0.9643 & 8.1566 & -1.2284 \\ 1.5858 & -3.8604 & -2.0710 & 3.2520 \\ 0.4633 & -1.4697 & -3.0432 & -7.2089 \end{pmatrix} \tag{24}$$

G refers to the Kalman filter gain, and B denotes the state feedback gain. Figure 14 shows that the LQG controller is capable of fast and accurate regulation of $\dot{m}_0$. By comparing Fig. 11 with Fig. 14, one can see that the LQG controller reduces the fuel oscillation peak by more than 600% within 100 ms.

6 Combustion Instability Control

Unstable combustion occurs at 274 Hz at an equivalence ratio of 0.62 and the power of 120 kW, roughly corresponding to the quarter wave mode of the combustion chamber. It is noticed that, the percentage of the pressure drop across the air swirler is about

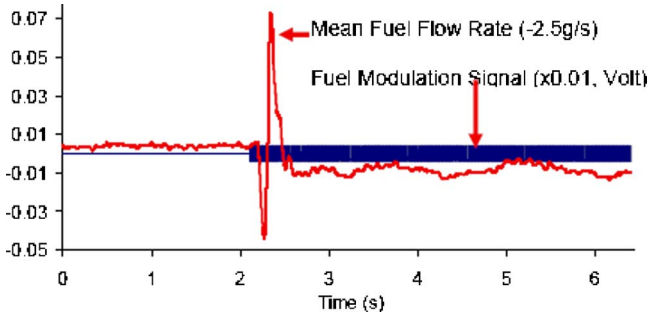


Fig. 14 LQG control of $\dot{m}_0$

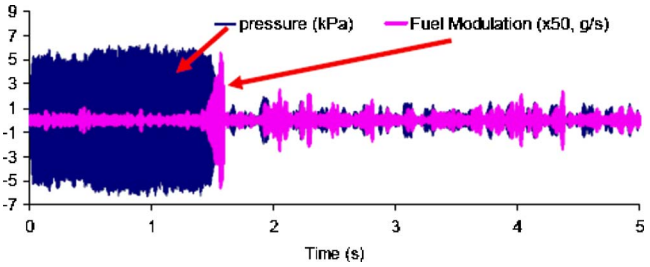


Fig. 15 Combustion instability control

30 times less than that across the fuel injectors. Thus, equivalence ratio variations, caused by the differential acoustic impedances between the fuel injectors and the air passage, may be the underlying mechanism for the occurrence of combustion instability. For the dynamics of the Titan oscillator to be improved, colored white noise with passband within [950–1000] Hz is added to the phase-shift control signal, which helps to keep the residual pressure oscillations at lower amplitude. Limited by space, these results are not included in the present paper.

Figure 15 shows that combustion instability is damped out within 130 ms. Spectral analysis shows that the unstable mode has been attenuated by more than 27 dB. Relatively large fuel modulations, about 5% of $\dot{m}_0$, are required at the initial phase of control. After that, a fuel modulation RMS less than 1.0% of $\dot{m}_0$ is able to prevent combustion instability from reoccurring. In this experiment, almost all the strategies favoring large fuel modulations have been implemented. Here L_1=1.75 m, L_2=0.33 m, A_{t1} =A_{t2}=7.1E−6 m^2, $\dot{m}_0$=3.5 g/s, and p_0=250 psi. The bypass fuel passage is used, and air has been removed from the fuel system at the beginning of the experiment. The LQG controller is working in parallel to the phase-shift controller. By comparing Fig. 12 with Fig. 15, we believe that the LQG controller has played an important role in this experiment. The small bursts in the residual pressure imply short-period loss of the control authority, which may be associated with the intrinsic inadequacies of the phase-shift controllers and the phase drift characteristics of combustion instability [14].

7 Conclusions

The present paper performs complementary theoretical and experimental investigations of a pump-style, high-frequency, magnetostrictive fuel actuator, which has potential applications for combustion instability control. Fuel modulations are achieved by "pushing" fuel out of a piston-cylinder unity using a Terfenol-D rod, and the mean fuel flow rate is controlled by a stepper motor. The mechanisms for fuel modulations and mean flow control are mechanically coupled. A system-identification-based pulse-width-modulation controller is designed to quickly suppress the strong interferences of fuel modulations with the mean fuel flow rate. Low-order models are developed to identify relevant parameters and investigate their influences on fuel modulations. Experiments validate the trend of the model predictions. Both the low-order models and experiments show that, large fuel modulations can be achieved with longer upstream tubing, shorter downstream tubing, a larger piston-cylinder unit, a bypass fuel passage, and fuel tubing with a larger diameter. Air or fuel bubbles within the fuel system significantly reduce fuel modulations and should be avoided as much as possible. It is also important for the fuel actuator to be of minimum-phase and to have fast dynamics, a small actuating dead zone, and minimum time delay. With the improved fuel setup and the LQG mean flow controller, effective control of combustion instability is demonstrated on an unstable gas turbine combustion simulator.

Acknowledgment

The authors would like to thank Mr. Michael Cornwell for designing and providing the high frequency fuel valve and assisting in the experiments.

Nomenclature

ρ = fuel density (about 860 kg/m^3 for turpentine), kg/m^3
η = fuel modulation ratio
ω = angular frequency, rad/s
ω_r = resonant frequency, rad/s
A_{t1} and A_{t2} = cross-sectional area of the upstream and the downstream tubing, respectively, m^2
A_c = cross-sectional area of the PCU, m^2
A_e = effective width of the exit ports of the PCU, m
K = isentropic modulus of fuel, Pa
L_1 = length of the upstream tubing, m
L_2 = length of the downstream tubing, m
$\dot{m}$ = mass flow rate of fuel, g/s
p_0 = pressure of the fuel supply, Pa
p_{atm} = the atmospheric pressure, Pa
p_c = pressure within the PCU, Pa
p_m = fuel pressure measured immediately upstream of the exit orifice, Pa
p_v = fuel pressure measured immediately downstream of the PCU, Pa
$\dot{Q}$ = volumetric flow rate of fuel, m^3/s
u_1 = bulk velocity within the upstream tubing, m/s
u_2 = bulk velocity within the downstream tubing, m/s
x_m = piston position (see Fig. 3), m
LQG = linear-quadratic-Gaussian
PCU = piston-cylinder unit
RMS = root-mean-square
$(\,')$ = perturbations of a variable
$(\,_0)$ = the mean value of a variable

References

[1] Lefebvre, A. H., 1995, "The Role of Fuel Preparation in Low-emission Combustion," ASME J. Eng. Gas Turbines Power, **117**, pp. 617–654.
[2] Gysling, D. L., Copeland, G. S., McCormick, D. C., and Proscia, W. M., 2000, "Combustion System Damping Augmentation With Helmholtz Resonators," ASME J. Eng. Gas Turbines Power, **122**, pp. 269–274.
[3] Eldredge, J. D., and Dowling, A. P., 2003, "The Absorption of Axial Acoustic Waves by a Perforated Liner With Bias Flow," J. Fluid Mech., **485**, pp. 307–335.
[4] Richards, G. A., Straub, D. L., and Robey, E. H., 2003, "Passive Control of Combustion Dynamics in Stationary Gas Turbines," J. Propul. Power, **19**(5), pp. 795–810.
[5] Neumeier, Y., Nabi, A., Arbel, A., Vertzberger, M., and Zinn, B. T., 1997, "Open-loop Performance of a Fast-Response Actively Controlled Fuel Injector Actuator," J. Propul. Power, **13**(6), pp. 705–713.
[6] Barooah, P., Anderson, T. J., and Cohen, J. M., 2003, "Active Combustion Instability Control With Spinning Valve Actuator," ASME J. Eng. Gas Turbines Power, **125**, pp. 925–932.
[7] Riley, A. J., Park, S., Dowling, A. P., Evesque, S., and Annaswamy, A. M., 2004, "Advanced Closed-Loop Control on an Atmospheric Gaseous Lean-Premixed Combustor," ASME J. Eng. Gas Turbines Power, **126**, pp. 708–716.
[8] Penner, S. S., 1955, "Similarity Analysis for Chemical Reactors and the Scaling of Liquid Fuel Rocket Engines," *Combustion Researches and Reviews*, AGARD Publications by Butterworths Scientific Publications.
[9] Munson, B. R.,Young, D. F., and Okiishi, T. H., 1990, *Fluid Mechanics*, Wiley, New York, 147 and 112.
[10] Ogata, K., 1997, *Modern Control Engineering*, 3rd ed., Prentice-Hall, Upper Saddle River, NJ, pp. 100–105 and 59.
[11] Coordinating Research Council, Inc, Atlanta, GA, 1983, Handbook of Aviation Fuel Properties, Report No. A601231.
[12] Nersesse, N., and Gregory, P. C., 2001, "Magneto-Thermo-Mechanical Characterization of Magnetostrictive Composites," Proc. SPIE, **4333**, pp. 166–177.
[13] Ljung, L., 1999, *System Identification: Theory for the User*, 2nd ed., Prentice-Hall, Upper Saddle River, NJ.
[14] Lieuwen, T., 2001, "Phase Drift Characteristics of Self-Excited, Combustion-Driven Oscillations," J. Sound Vib., **242**(5), pp. 893–905.

Experimental and Kinetic Modeling of Kerosene-Type Fuels at Gas Turbine Operating Conditions

P. Gokulakrishnan[1]
e-mail: gokul@csefire.com

G. Gaines

J. Currano[2]

M. S. Klassen

R. J. Roby

Combustion Science & Engineering, Inc.,
8940 Old Annapolis Road, Suite L,
Columbia, MD 21045

Experimental and kinetic modeling of kerosene-type fuels is reported in the present work with special emphasis on the low-temperature oxidation phenomenon relevant to gas turbine premixing conditions. Experiments were performed in an atmospheric pressure, tubular flow reactor to measure ignition delay time of kerosene (fuel–oil No. 1) in order to study the premature autoignition of liquid fuels at gas turbine premixing conditions. The experimental results indicate that the ignition delay time decreases exponentially with the equivalence ratio at fuel-lean conditions. However, for very high equivalence ratios (>2), the ignition delay time approaches an asymptotic value. Equivalence ratio fluctuations in the premixer can create conditions conducive for autoignition of fuel in the premixer, as the gas turbines generally operate under lean conditions during premixed prevaporized combustion. Ignition delay time measurements of stoichiometric fuel–oil No. 1/air mixture at 1 atm were comparable with that of kerosene type Jet-A fuel available in the literature. A detailed kerosene mechanism with approximately 1400 reactions of 550 species is developed using a surrogate mixture of n-decane, n-propylcyclohexane, n-propylbenzene, and decene to represent the major chemical constituents of kerosene, namely n-alkanes, cyclo-alkanes, aromatics, and olefins, respectively. As the major portion of kerosene-type fuels consists of alkanes, which are relatively more reactive at low temperatures, a detailed kinetic mechanism is developed for n-decane oxidation including low temperature reaction kinetics. With the objective of achieving a more comprehensive kinetic model for n-decane, the mechanism is validated against target data for a wide range of experimental conditions available in the literature. The data include shock tube ignition delay time measurements, jet-stirred reactor reactivity profiles, and plug-flow reactor species time–history profiles. The kerosene model predictions agree fairly well with the ignition delay time measurements obtained in the present work as well as the data available in the literature for Jet A. The kerosene model was able to reproduce the low-temperature preignition reactivity profile of JP-8 obtained in a flow reactor at 12 atm. Also, the kerosene mechanism predicts the species reactivity profiles of Jet A-1 obtained in a jet-stirred reactor fairly well. [DOI: 10.1115/1.2436575]

Keywords: gas turbines, lean premixed combustion, kerosene oxidation, kinetic modeling, flow reactor, autoignition, aviation fuels

Introduction

In recent years, stringent emission standards have made lean-premixed combustion more popular in industrial applications than ever before. However, a difficulty often encountered in practical combustion devices is the premature auto-ignition of fuel/air mixture before it reaches the main combustion zone. This is especially true with the higher order hydrocarbon fuels, such as kerosene and diesel, which have a shorter ignition delay time compared to natural gas [1]. The emergence of liquid prevaporized lean-premixed combustion in gas turbines [2] and fuel reforming to produce hydrogen-enriched gas for fuel cells via cool flame partial oxidation [3] has revitalized the importance of low-temperature hydrocarbon oxidation, which has been the subject of knocking phenomenon in internal combustion engines for the last 3 decades [4].

Kerosene is a generic term used for a range of fuels from fuel oil No. 1 to more refined aviation grade fuels. Because of its storage and thermal stabilities, kerosene-type aviation fuels are used in aircraft gas turbines engines. Notably, kerosene-type Jet A and Jet A-1 are widely used, respectively, in domestic and international civilian passenger jets, while JP-5 and JP-8 are used by the US Navy and US Air Force, respectively [5]. ASTM chemical analysis of kerosene used in the present work shows that the fuel contains 77.6 vol % alkanes, 20.0 vol % aromatics, and 2.4 vol % olefins, and negligible amount of naphthenes (i.e., 0.4 wt %). On the other hand, kerosene-type aviation grade fuels, for example JP-8, consist of 60 vol % alkanes, 18 vol % aromatics, 20 vol % naphthenes, and 2 vol % olefins on average [6] though composition may vary widely from sample to sample. The presence of a significant amount of naphthenes enhances the thermal stability of the jet fuels. Because the naphthenes undergo endothermic fuel decomposition at low temperatures, the kerosene can act as a heat sink for the cooling of the jet engines [7]. It should also be noted

[1]Corresponding author.
[2]Current address: Department of Mechanical Engineering, University of Maryland, College Park, MD 20742.

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received December 22, 2005; final manuscript received May 31, 2006. Review conducted by Dilip R Ballal. Paper presented at the ASME Turbo Expo 2005: Land, Sea and Air (GT2005), Reno, NV, June 6–9, 2005. Paper No. GT2005-68268.

Table 1 Surrogate mixtures available in the literature to model Kerosene Oxidation: Surrogate components of kerosene

Components (vol %)		Dagaut[a]	Violi et al.[b]	Lindstedt and Maurice[c]
n-propyl-benzene	C_9H_{12}	15		
n-propyl-cyclohexane	C_9H_{18}	11		
n-decane	*n*-$C_{10}H_{22}$	74	25	89
iso-octane	*i*-C_8H_{18}		5	
clyclo-hexane	C_7H_{14}		5	
toluene	C_7H_8		20	
dodecane	*n*-$C_{12}H_{26}$		25	
teradecane	*n*-$C_{14}H_{30}$		20	
aromatics				11

[a]See Ref. 14.
[b]See Ref. 15.
[c]See Ref. 17.

that additional chemical enhancers have been used in the aviation fuels to increase the storage and thermal stability of the fuels, for example, JP-8 +100 [8].

Each hydrocarbon group identified above is a mixture of a wide range of individual hydrocarbon components, but they may not necessarily share common oxidation characteristics, except *n*-alkanes with the number of carbon atoms greater than four. Roubaud et al. [9] performed a comparative experimental study in a rapid compression machine (RCM) to test the auto-ignition behavior of different types of alkyl-benzenes at 900 K varying the pressure. It was observed that, for example, toluene, *m*-xylene, and *p*-xylene could not auto-ignite below 16 bar, while *o*-xylene and *n*-butylbenzene were able to auto-ignite at 10 bar in 50 ms [9]. This illustrates the fact that the type of aromatics present in kerosene will influence the auto-ignition at low temperatures. It should also be noted that the ranking of the reactivity of the hydrocarbon groups mentioned above is: *n*-alkanes > cyclo-alkanes > aromatics [10]. Hence, the low-temperature oxidation is largely determined by the *n*-alkanes, as cylo-alkanes and aromatic components are less reactive compared to *n*-alkanes. Therefore, *n*-decane is often used as a primary surrogate fuel to study practical fuels such as kerosene [11] and diesel [12].

Agosta et al. [13] performed a series of high-pressure, plug-flow reactivity experiments at pre-ignition conditions to determine a chemical surrogate that is comparable to JP-8 fuel (60% paraffins, 20% naphthenes, 18% aromatics, and 2% olefins). The closest surrogate mixture that was able to reproduce the reactivity profile of JP-8 was 26% *n*-dodecane, 34% iso-cetane, 14% methylcyclohexane, 6% decalin, and 18% α-methylnaphthalene. However, it would be an arduous task to develop and implement such a kinetic model for this five-component surrogate mixture to predict JP-8 oxidation. Thus, in order to simplify the complexity of fuel mixtures for modeling purposes, chemical surrogates are devised to represent major hydrocarbon groups with fairly well-studied components. A successful model depends on how well the model predicts oxidation phenomenon of its individual surrogate member. Table 1 shows three different types of chemical surrogates used in the literature to model JP-8 type aviation grade fuels. Dagaut [14] demonstrated that the high temperature reactivity profile of Jet A-1 can be modeled using a surrogate mixture of 74% *n*-decane, 15% *n*-propylbenzene, and 11% *n*-propylcylohexane representing *n*-alkanes, aromatics and cyclohexanes, respectively, while Violi et al. [15] used a six-component surrogate mixture to model JP-8, representing three types of *n*-alkanes along with iso-octane, methylcyclohexane, and toluene. Mawid et al. [16] used different surrogate mixtures with the number of chemical constituents as high as 12 to model kerosene oxidation. Lindstedt and Maurice [17] on the other hand just considered *n*-decane and a mixture of aromatics.

One of the challenging aspects of kinetic modeling of any petroleum-derived fuel is to predict the oxidation in low to intermediate temperatures (<1000 K), in which alkanes undergo three distinct regimes of combustion, namely slow combustion, cool flame, and negative temperature coefficient (NTC) region [10]. As noted in Table 1, *n*-decane has been used as a primary surrogate to model kerosene oxidation. However, the modeling tasks of the mechanisms noted in Table 1 were focused on high-temperature kerosene oxidation. Thus, it is crucial for the *n*-decane mechanism to be validated against low-temperature experimental data, especially under lean conditions between 600 and 800 K, to make predictions at gas turbines premixing conditions. Two *n*-decane mechanisms with low-temperature chemistry were reported in the literature [18,19]. The mechanism of Bikas and Peters [18], consisting of lumped low-temperature *n*-decane kinetics, was not validated for a low-temperature species–time history profile. However, the model was validated against shock tube ignition delay time of Pfahl et al. [12] for stoichiometric or higher equivalence ratios. On the other hand, the mechanism of Battin-Leclerc et al. [19], consisting of 4800 reactions and 1100 species, was validated for reactivity and species time–history measurements. Thus, in this work, a detailed *n*-decane mechanism is developed by validating against ignition delay time, reactivity profiles, and species time–history profiles over a wide range of temperatures and pressures to represent *n*-alkanes in the kerosene mechanism. Furthermore, in the kerosene mechanism developed in the present work, *n*-alkanes, aromatics, and naphthenes are represented by *n*-decane, *n*-propylbenzene, and *n*-propylcylohexane, respectively. Olefins are represented by decene ($C_{10}H_{20}$), which is a reaction subset in the *n*-decane mechanism. The kinetic mechanisms for *n*-propylbenzene and *n*-propylcylohexane are adopted from Dagaut [14].

Kinetic Model Development and Validation

During the oxidation of hydrocarbons, the fuel molecules first undergo H-atom abstraction via reaction (R1) to produce alkyl radicals

$$RH + X = R + XH \tag{R1}$$

where X represents species OH, H, O, HO_2, O_2, or CH_3. In high-temperature oxidation, the alkyl radicals further decompose to produce smaller olefins such as ethylene and propene by β-scission reactions, while the runaway reaction (R2) dominates the chain branching step to generate hydro-oxy radicals such as O, H, OH, and HO_2

$$H + O_2 = O + OH \tag{R2}$$

However, at low temperatures, the recombination reaction of H with O_2 dominates the consumption of H to yield relative unreactive HO_2 radicals via reaction (R3)

$$H + O_2 + M = HO_2 + M \tag{R3}$$

Moreover, the alkyl radicals produced in reaction (R1) by H-atom abstraction go through more complex propagation and chain branching steps, in which the alkyl radicals first react with molecular O_2 to form RO_2 radical in the absence of reaction (R2)

$$R + O_2 = RO_2 \tag{R4}$$

The RO_2 radical, formed in reaction (R4), would then proceed through internal H-atom abstractions and isomerization reactions to produce smaller olefins via intermediate oxygenated heterocyclic rings, while generating OH radicals in the process.

Although the forward reaction (R4) can readily occur at low temperatures, equilibrium shifts in the reverse direction as the temperature increases to a certain level known as the turnover temperature. This turnover temperature indicates the beginning of an NTC region [20]. The transition from low to intermediate temperature regime is defined as the temperature at which the termination of the NTC occurs. In reaction (R4), the equilibrium shifts in the reverse direction so that the low-temperature reaction route

starts to diminish as the temperature is increased above the turnover temperature. When the temperature reaches the end of the NTC regime, the reverse reaction (R4) is favored and beyond this point the oxidation proceeds via an intermediate temperature route. It is noteworthy that, as the pressure increases, the turnover temperature (the temperature at which NTC begins) shifts toward higher values, and the NTC region becomes narrower [20]. Thus, the definition of low- to intermediate-temperature regimes can vary with pressure during alkane oxidation.

As the primary focus of most of the studies reported in the literature has been on high temperature oxidation of *n*-decane [17,18,21–23], a new low-temperature kinetic mechanism is developed in this study as a basis to build a kerosene mechanism having *n*-propylcyclohexane and *n*-propylbenzene to represent naphthenes and aromatics, respectively. Ignition delay time measurements in an RCM [24] found that cyclohexane shows cool flame and NTC regions. Upon opening the cyclic ring, the hexenyl radical undergoes low-temperature isomerization steps at low temperatures similar to *n*-alkanes [25]. Also, it is possible that the alkyl side chains attached in alkyl-cycloparaffins and alkyl-aromatics can go through similar kinetic steps as in *n*-alkane low-temperature oxidation. In the present mechanism, the alkyl side chains are represented by the propyl radical, and hence a detailed propane oxidation mechanism is included. The reaction rate coefficients for high-temperature *n*-propylbenzene and *n*-propylcylohexane oxidation are adapted from Dagaut [14]. Also, a low-temperature reaction submechanism for the parent ring in the *n*-propylcylohexane is included in the present kerosene mechanism. The reaction path and the rate parameters are adapted from the recommendations of Granata et al. [25].

A new detailed *n*-decane kinetic mechanism is developed following the steps that Curran et al. [26] used in developing an *n*-heptane mechanism. Due to space constraints, the main steps of the model development are described here. At low temperatures, *n*-decane molecules undergo H-atom abstraction (six primary and 16 secondary H-atom abstraction) via reaction (R1) to produce five possible decyl radicals [22], which would then individually react with molecular O_2 to form RO_2 following reaction (R4). The rate coefficients of H-atom abstraction are estimated from the additivity rule [27] using the values recommended by Curran et al. [26], as the pre-exponential factor and the activation energy per C–H bond becomes constant for alkanes with higher than four carbon atoms [28]. Each RO_2 radical then undergoes unimolecular, internal H-atom abstraction, and isomerization to produce QOOH forming a heterocyclic transition state ring

$$RO_2 = QOOH \tag{R5}$$

However, the reactions involving five to eight-member transition state rings were only considered assuming that the seven-member transition rings have the lowest activation energy [29]. Hence, only 30 QOOH isomeric radicals are considered. The rate coefficients are estimated by the additivity rule [27] from the values recommended for the corresponding *n*-heptane counterparts [26]. Reaction (R5) becomes one of the main rate-limiting steps for the modeling of low-temperature ignition delay time measurements reported by Pfahl et al. [12]. Thus, the pre-exponential factor for RO_2 internal isomerization reactions, recommended by Curran et al. [26] for *n*-heptane, was increased by an order of magnitude to increase the reactivity of *n*-decane decomposition during low-temperature oxidation.

QOOH then proceeds to take two different routes: propagation and chain branching. During propagation, decomposition of QOOH generates either the OH or HO_2 radical while forming conjugate olefins (R6), cyclic ether (QO) (R7), and smaller olefins via β-scission (R8)

$$QOOH \Rightarrow \text{conjugate-olefins} + HO_2 \tag{R6}$$

$$QOOH \Rightarrow QO + OH \tag{R7}$$

$$QOOH \Rightarrow \beta\text{-scission products} + OH \tag{R8}$$

In a chain branching step, QOOH reacts with molecular O_2 to form O_2QOOH, which then isomerizes to form ketohydroperoxide via reaction (R10)

$$QOOH + O_2 = O_2QOOH \tag{R9}$$

$$O_2QOOH \Rightarrow O_2QO + OH \tag{R10}$$

The O_2QOOH radicals formed in reaction (R9) undergo internal H-atom abstraction isomerization to form ketohydroperoxide (O_2QO) and OH. The ketohydroperoxides decompose to generate oxygenated radical species and OH. Thus, it should be noticed that the route $RO_2 \Rightarrow QOOH \Rightarrow O_2QOOH \Rightarrow O_2QO$ eventually becomes a branching step of the low-temperature oxidation as it generates two OH radicals along with smaller olefins and oxygenated species such as aldehydes.

At high temperatures (>1000 K), the reverse reaction (R4) is favored, and hence, the decyl radicals formed in reaction (R1) undergo β-scission to form smaller olefins and alkyl radicals via reaction (R11)

$$C_{10}H_{21}-1 = C_8H_{17}-1 + C_2H_4 \tag{R11a}$$

$$C_{10}H_{21}-2 = C_7H_{15}-1 + C_3H_6 \tag{R11b}$$

$$C_{10}H_{21}-3 = C_6H_{13}-1 + C_4H_8-1 \tag{R11c}$$

$$C_{10}H_{21}-4 = C_5H_{11}-1 + C_5H_{10}-1 \tag{R11d}$$

$$C_{10}H_{21}-5 = pC_4H_9 + C_6H_{12}-1 \tag{R11e}$$

In addition, *n*-decane undergoes unimolecular decomposition to form smaller alkyl radicals via C–C homolysis reactions at high temperatures [22]

$$nC_{10}H_{22} = C_9H_{19}-1 + CH_3 \tag{R12a}$$

$$nC_{10}H_{22} = C_8H_{17}-1 + C_2H_5 \tag{R12b}$$

$$nC_{10}H_{22} = C_7H_{15}-1 + nC_3H_7 \tag{R12c}$$

$$nC_{10}H_{22} = C_6H_{13}-1 + pC_4H_9 \tag{R12d}$$

$$nC_{10}H_{22} = C_5H_{11}-1 + C_5H_{11}-1 \tag{R12e}$$

Also, the chain branching reaction (R2) becomes the dominant route for the generation of hydro-oxy radical pool at high temperatures. The subsequence reaction paths for the olefins and alkyl radicals with less than seven carbon atoms, produced in reactions (R11) and (R12), are adapted from the *n*-heptane mechanism of Curran et al. [26]. The rate coefficients for the reactions involving species with higher than seven carbon atoms are taken from the literature on high temperature *n*-decane oxidation [14,22].

A sensitivity analysis [30] was performed for different experimental conditions in order to identify the rate-limiting steps at different temperatures and pressures using the Senkin code [31]. This information is then used to optimize the reaction mechanism by tuning the rate coefficient of relevant reactions within its uncertainty to match the model predictions with the experimental measurements. Figures 1, 2, and 3 show the normalized sensitivity coefficients (NSC) for the ten most important rate-limiting reactions with respect to *n*-decane at 650, 950, and 1250 K, respectively. The temperatures were chosen to represent different regimes of combustion, namely the low, intermediate, and high temperatures. The sensitivity coefficients were computed for a stoichiometric *n*-decane/air mixture at 13 atm pressure, similar to the shock tube experimental conditions of Pfahl et al. [12]. It should be noted that, in Fig. 1, species O_2QOOH and O_2QO (ketohydroperoxides) are denoted by $C_{10}H_{20}OOHx-yO_2$ and $nC_{10}ketxy$, respectively, where *x* and *y* refer to the carbon number at which an O atom is attached. As demonstrated in Fig. 1 at 650 K, the for-

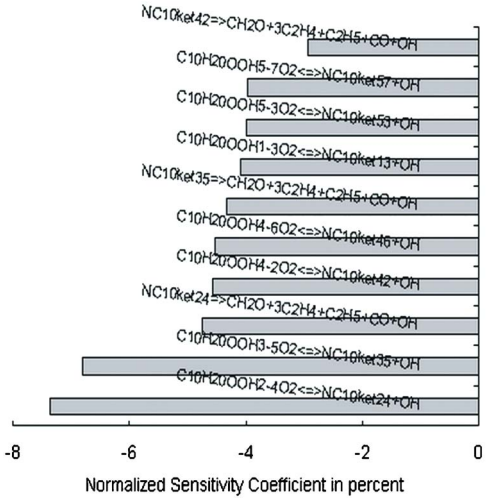


Fig. 1 Ten reactions with the highest normalized sensitivity coefficients (NSC) with respect to *n*-decane for stoichiometric *n*-decane/air mixture at 13 atm. The initial temperature was 650 K.

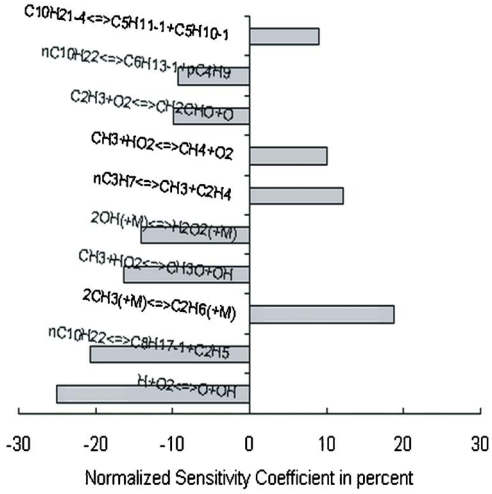


Fig. 3 Ten reactions with the highest NSC with respect to *n*-decane for the conditions specified in Fig. 1 at 1250 K

mation of ketohydroperoxides (i.e., reaction (R10)) and their subsequent decomposition (i.e, reactions NC10ket$xy \Rightarrow CH_2O + 3C_2H_4 + C_2H_5 + CO + OH$) become the most sensitive reactions at low temperatures. Thus, according to the sensitivity analysis illustrated in Fig. 1, the chain branching sequence, $RO_2 \Rightarrow QOOH \Rightarrow O_2QOOH \Rightarrow O_2QO$, dominates the *n*-decane oxidation at low temperatures. However, in the intermediate temperature regime, reactions involving H_2O_2 become the predominant rate-limiting steps along with decyl β-scission reactions as noted in Fig. 2. It is interesting to note that the H-atom abstraction by HO_2 radical becomes the most sensitive among the decyl radical forming reactions at 950 K. However, H-atom abstractions by OH radicals are the most sensitive abstraction reactions at low temperatures. This is due to the fact that the relatively unreactive HO_2 radical formed in reaction (R3) becomes more reactive in the intermediate temperatures by the formation of H_2O_2 via reaction (R13)

$$2HO_2 = H_2O_2 + O_2 \quad \text{(R13)}$$

Then, the H_2O_2 dissociates to form OH radicals via reaction (R14)

$$H_2O_2 + (M) = 2OH(+M) \quad \text{(R14)}$$

It is noteworthy that dominance of H_2O_2 signifies the end of the NTC regime as the hydro-oxy radical pool is increased via reaction (R14) [20].

The sensitivity analysis depicted in Fig. 3 shows the most important reactions during high temperature oxidation of *n*-decane. It can be observed that the chain branching step (R2) is the most sensitive reaction of all. It is also interesting to note that C–C homolysis reactions (R12*b*) and (R12*d*), and β-scission reaction (R11*d*) are the most significant reactions involving *n*-decane or decyl radicals among the top ten reactions shown in Fig. 3.

The kerosene mechanism developed in this work is validated for *n*-decane using the shock tube ignition delay time measurements [12], species time–history profile obtained in plug-flow reactor [22], and reactivity profile obtained jet-stirred reactor [14] at various experimental conditions. Zero-dimensional simulations are performed using the Senkin code [31] assuming constant volume approximation for the ignition delay time computations under reflected shock wave conditions. The Sandia perfectly stirred reactor (PSR) code [32] was used for stirred reactor simulations.

Figures 4, 5, and 6 compare the model predictions against

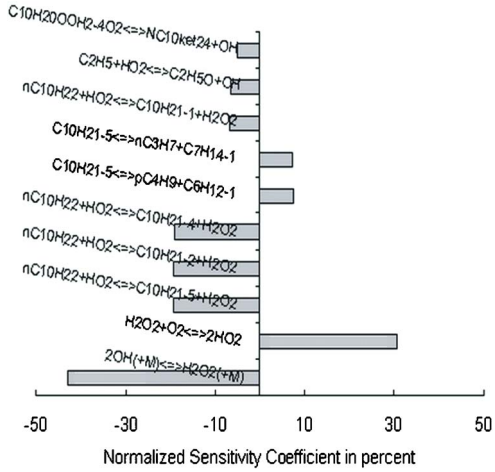


Fig. 2 Ten reactions with the highest NSC with respect to *n*-decane for the conditions specified in Fig. 1 at 950 K

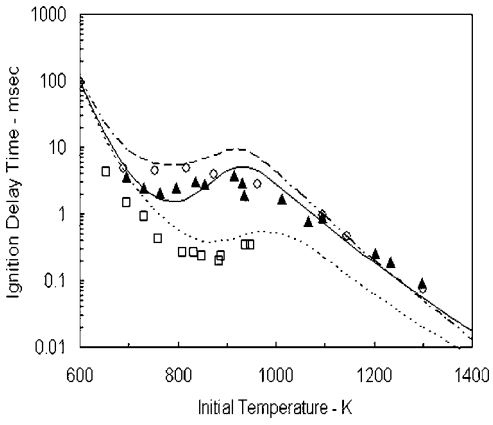


Fig. 4 Comparison of *n*-decane ignition delay time model predictions with shock tube measurements of Pfahl et al. [12] at various experimental conditions as indicated. The symbols and the lines denote the experimental measurements and model predictions, respectively: (○)—φ=0.5 at 13 atm; (▲)—φ=1.0 at 13 atm; (□)—φ=1.0 at 50 atm (φ—equivalence ratio).

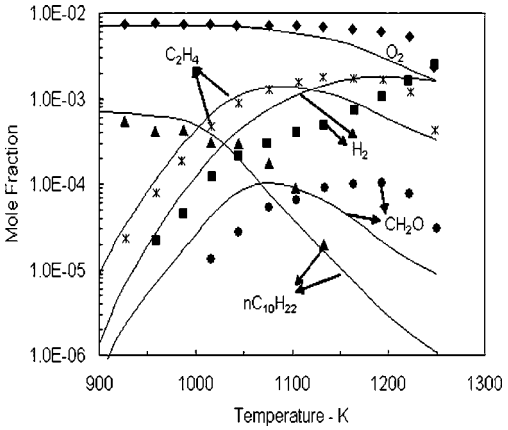


Fig. 5 Model comparison for *n*-decane reactivity species profile from jet-stirred reactor experiments of Dagaut [14] performed at 1.5 equivalence ratio and 1 atm pressure. The jet-stirred reactor residence time was 70 ms. The symbols and the lines denote the experimental measurements and model predictions, respectively: (●) CH_2O; (▲) $nC_{10}H_{22}$, (■) H_2; (♦) O_2, (∗) C_2H_4.

n-decane experimental measurements obtained in shock tube, jet-stirred reactor, and plug-flow reactor, respectively. Figure 4 shows the ignition delay time measurements of Pfahl et al. [12] measured behind reflected shock waves for *n*-decane/air mixtures at 13 atm and 50 atm pressures and 0.5 and 1.0 equivalence ratios. In Fig. 4, the model predictions show a very good agreement at 13 atm pressure. It is noteworthy that the model correctly identifies the NTC region, although the modeling results show a slight overprediction at 50 atm pressure. It should be noted that the low-temperature reaction subset was tuned with respect to most sensitive reactions with large uncertainty on their rate coefficients in order to optimize the ignition delay time predictions at different pressures. As recommended by Curran et al. [26], the pre-exponential factor was reduced by a factor of two and the activation energy was decreased by 3 kcal/mol for the isomerization of O2QOOH to form ketohydroperoxides and OH. In addition, the pre-exponential factor for ketohydroperoxide (O2QO) decomposition reaction was also increased by an order of magnitude.

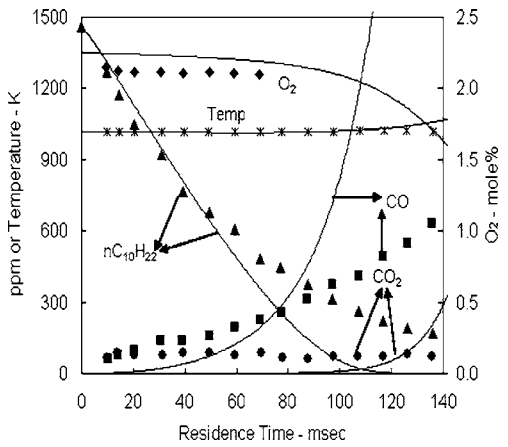


Fig. 6 Model comparison for *n*-decane species time–history profile from plug-flow experiments of Zeppieri et al. [21] performed at 1019 K and 1 atm. The symbols and the lines denote the experimental measurements and model predictions, respectively: (●) CO_2; (■) CO, (▲) $nC_{10}H_{22}$, (♦) O_2, (∗) temperature.

Figure 5 shows the model predictions for the reactivity of the diluted mixture of *n*-decane/O_2/N_2 in an atmospheric, jet-stirred reactor between 900 K and 1250 K [14]. Dagaut [14] measured the species profile varying the reactor temperature from 900 K to 1250 K, while the residence time of the reactor and the equivalence ratio of the inlet feed were maintained at 70 ms and 1.5, respectively. The inlet to the reactor contains 700 ppm of *n*-decane. On the other hand, Fig. 6 shows the species time–history profile of diluted, stoichiometric mixture of *n*-decane/O_2/N_2 oxidation obtained in an adiabatic plug-flow reactor at 1 atm and 1019 K. The composition of *n*-decane in the initial mixture was 1450 ppm. In both cases, the model predicts the *n*-decane decomposition profile reasonably well. However, the model inherits some deficiency in predicting intermediate species profiles, although in most instances the model predicts the experimental trends correctly. For example, Fig. 6 shows the model predictions for CO were higher than the experimental measurements above 100 ms residence time.

Experimental Apparatus

To develop a reliable kinetic mechanism for hydrocarbon oxidation, the rate parameters of the elementary reactions must be known fairly accurately. However, there is a considerable uncertainty in the estimation of these rate parameters due to lack of experimental measurements. Thus, validation of kinetic models is very often performed against ignition delay time, and reactivity and species time–history profiles available in the literature. Shock tube and RCM have been widely used to measure ignition delay time measurements at spark ignition engine relevant conditions. Plug-flow reactors have been used to measure reactivity as well as species time–history profile for hydrocarbon fuels. Moreover, Freeman and Lefebvre [33] and Spadaccini and TeVelde [34] used flow reactors to measure the ignition delay time of aviation grade fuels at gas turbine relevant conditions.

Since there is a scarcity of experimental measurements of the autoignition behavior of "real" fuels such as kerosene and diesel, a flow reactor is used in the present work to measure the ignition delay time at conditions relevant to premixing of fuel and air in gas turbines. The experiments were performed in an atmospheric pressure flow reactor using heptane and kerosene (i.e., fuel oil No. 1). A schematic of the experimental system is shown in Fig. 7. The flow reactor consists of three parts: (1) liquid fuel vaporizing section; (2) fuel–air premixing section; and (3) test section. The liquid fuels are prevaporized in the vaporizing section before they are premixed with air in the premixing section. Then, the premixed gas mixture travels through the test section where ignition occurs. Figures 8 shows the schematic of the premixing section.

The vaporizer was maintained at 300°C by flowing preheated nitrogen through the vessel as the carrier gas. The liquid fuel is continuously injected into the vaporized vessel via a gas-blast pressure atomizer using nitrogen as the atomizing gas. The total molar flow of N_2 to fuel was maintained at 10:1 to prevent condensation of fuel vapor on its passage to the premixer. The down stream of the vaporizer is connected to two solenoid valves; one is normally open to the exhaust, while the other one is connected to the premixer. The fuel vapor is passed through the open solenoid valve to the exhaust continuously. When the solenoid valves are activated, the solenoid valve open to the exhaust is closed and the second valve allows the fuel vapor into the premixer. Then, the fuel vapor mixes with air in the premixing section and then travels through the test section, where a photodiode records the chemiluminescence emissions. The time between the solenoid valve activation and the light emission is defined as the ignition delay time.

As most of the data from flow reactors are used to validate kinetic models (due to the simplified fluid mechanics of these systems), it is very important to minimize the mixing effects of the fuel and air at the entrance to the flow reactor so that the numerical simulation of an "ideal" plug-flow reactor can be compared to the experimental data. However, in reality, even rapid

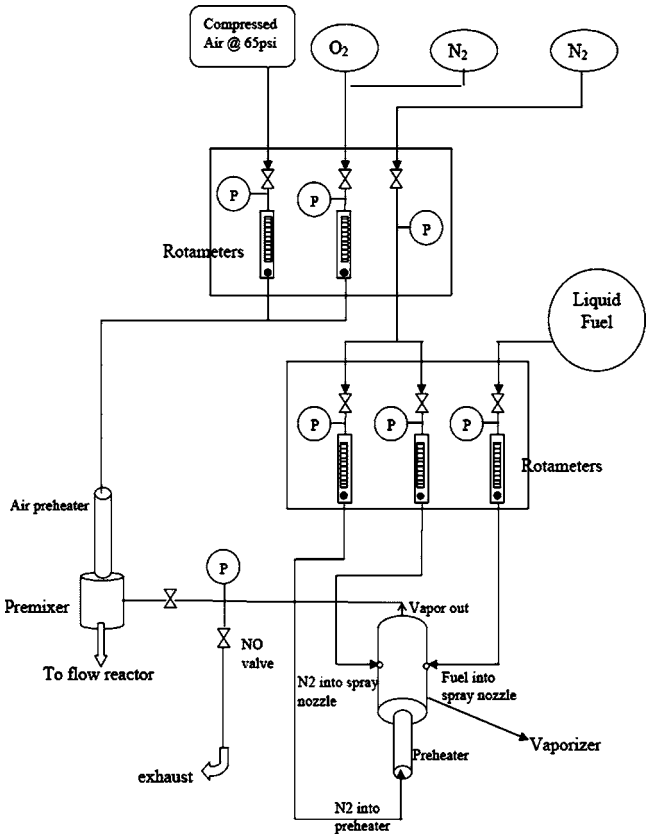


Fig. 7 Schematic of the experimental system

mixing of fuel and air has a finite time scale for the mixing. Yetter et al. [35] and Gokulakrishnan et al. [36] have demonstrated the treatment of mixing effects during numerical simulation of plug-flow reactor experiments. In the present flow reactor setup, a swirler is employed in the premixing section to achieve rapid mixing of fuel and air before the mixture is sent into the test section. As shown in Fig. 8, the upstream portion of the premixing section consists of a swirler to enhance the mixing of the fuel and air by creating turbulence eddies. An optimized design configuration of the premixing section was obtained from a fluid mechanical simulation of the system with a tracer demonstrating the mixing behavior of fuel and air.

The downstream portion of the premixing section consists of an expanding duct in order to avoid flow separation while increasing the mixing of the incoming flow. The downstream of the expanding duct is a (2 in. diameter $\times$ 52 in. long) alumina (Al_2O_3) tube

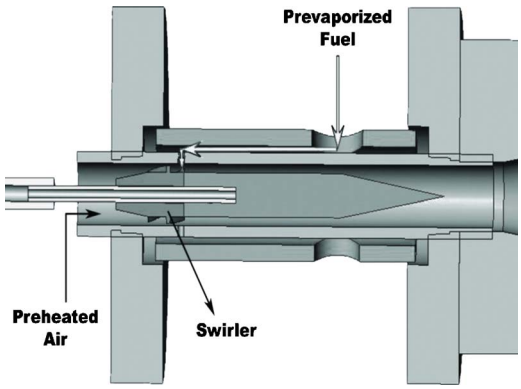


Fig. 8 Schematic of the premixing section of the flow reactor

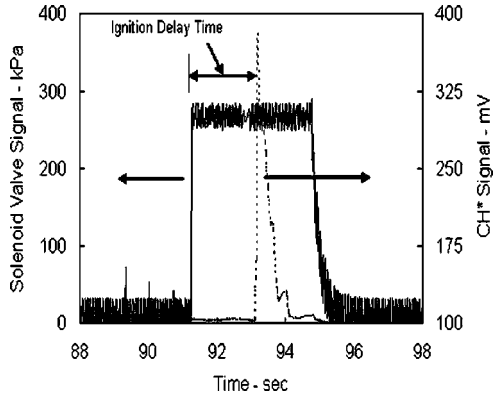


Fig. 9 Typical experimental data recorded during the ignition delay time measurements: (solid line) solenoid valve signal; (dotted line) CH^* emission signal detected by photomultiplier coupled with CH narrow band filter

which acts as the test section of the flow reactor. The test section is placed in a tubular furnace with independently controlled three-zone heaters. As the heaters are placed in a well insulated ceramic box, the test section can be assumed to be at adiabatic conditions. The plug-flow test section is maintained at uniform initial reactor temperature with ±5 K temperature variation. The swirler and expanding duct help to meet the pre-requisites for ideal plug-flow conditions in the test section, in which no micromixing would occur along the axis of reactor. Also, rapid mixing of air and fuel at the entrance to the flow reactor is very important in order to reduce the error introduced by the "unmixedness" of the reactants in the ignition delay time measurements. An amplified photodiode coupled with a CH^* narrow band filter is placed at the end of the test section to detect CH^* emissions. A three-way solenoid valve, placed upstream of the premixer as shown in Fig. 7, would let the fuel vapor into the premixer, when the valve is activated. In normal mode, the valve only let the fuel vapor into the exhaust. The time difference between the injection of the prevaporized fuel into the mixing section and the detection of CH^* emission signal is defined as the ignition delay time. Figure 9 shows a typical recording of the solenoid valve opening and the detection of CH^* signal in the photodiode.

Results and Discussion

Initially, *n*-heptane was used as prototype fuel to measure the ignition delay time in the flow reactor so that results can be compared with other experimental systems, as *n*-heptane is the most widely studied higher-order hydrocarbon [26]. The ignition delay time obtained in the flow reactor is compared with the shock tube ignition delay time measurements available in the literature. The ignition delay time of stoichiometric *n*-heptane/air mixture was measured varying the flow reactor temperature between 700 K and 1000 K. Figure 10 compares the *n*-heptane ignition delay time measured in the present study with the shock tube measurements obtained at 1 atm [37] and 13 atm [38] pressures. Also, the experimental measurements are compared with the model predictions of Curran et al. comprehensive *n*-heptane mechanism [26] that was validated against a wide range of experimental measurements. As seen in Fig. 10, the model predictions agree with the experimental measurements very well, both qualitatively and quantitatively.

Figure 11 shows the ignition delay time measurements of kerosene (fuel oil No. 1) and *n*-heptane as a function of equivalence ratio. The kerosene ignition delay time was measured at different equivalence ratios by keeping the initial O_2 level at 17% and the temperature at 900 K. The ignition delay time measurements of *n*-heptane were made at 865 K and 10% O_2. The experimental measurements are compared with the model predictions of kero-

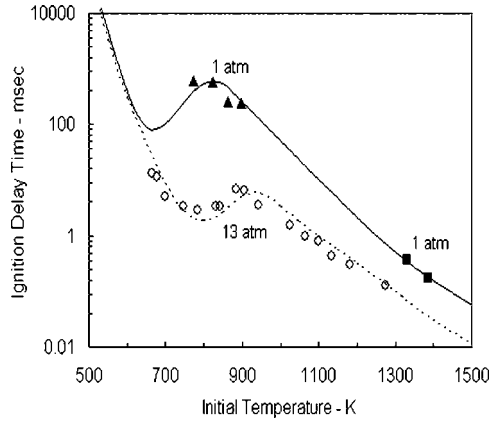


Fig. 10 Flow reactor ignition delay time measurements of stoichiometric *n*-heptane/air mixtures compared with shock tube ignition delay time measurements as well as model predictions of Curran et al. [25]; (▲) present work, (■) Horning [33], and (○) Ciezki et al. [34]. The lines indicate the modeling results of Curran et al. [25].

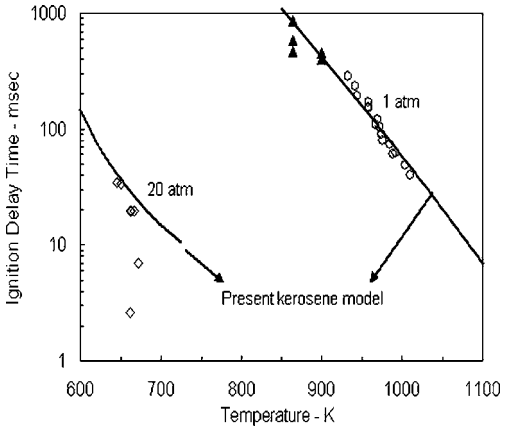


Fig. 12 Jet A/air ignition delay time measurements of Freeman and Lefebvre and [31] (denoted by ◇) and Spadaccini and TeVelde [32] (denoted by ○) at 1 atm and 20 atm pressures, respectively. The solid symbols (▲) indicate the ignition delay time of kerosene/air obtained in the present study at 0.50–0.75 equivalence ratio. The lines denote the kerosene model predictions.

sene mechanism developed in the present work as well as the *n*-heptane mechanism of Curran et al. [26]. The model predictions shown in Fig. 11 agree fairly well with experimental measurements.

It can be observed that ignition delay time increases exponentially with decreasing equivalence ratios at fuel-lean conditions for *n*-heptane and fuel oil No. 1 at 865 K and 900 K, respectively. However, the ignition delay time of fuel-rich mixtures tends to approach an asymptotic as the equivalence ratio is increased. These results show the importance of the impact of equivalence ratio fluctuations in determining the ignition delay time at premixing conditions of a gas turbine where the fluctuation of the equivalence ratio is so large that pockets of fuel-rich mixtures can trigger the ignition prematurely. It is noteworthy that this phenomenon, i.e., the leaner the mixture the longer the ignition delay time, is only true at low to intermediate temperatures. At higher temperatures, the leaner mixtures have shorter ignition delay times. For example, in Fig. 4, *n*-decane ignition delay time for 0.5 equivalence ratio at 13 atm remains higher than the stoichiometric mixture at the same conditions until 1200 K. At temperatures higher than 1200 K, the ignition delay time of lean mixture is lower than that of the stoichiometric *n*-decane/air mixture. This is due to the competition between fuel molecules and O_2 for H atoms at high temperatures [39]. Thus, the increase in fuel concentration inhibits the availability of H for chain branching step (R2), which results in an increase in the ignition delay time at higher equivalence ratios as observed in Fig. 4 above 1200 K. However, at low temperatures, the main oxidation occurs through the formation of alkylperoxides via reaction (R3) and subsequent low temperature chain branching steps described in Fig. 1. Therefore, fuel-rich mixtures have the advantage of generating alkyl radicals at lower temperatures, and hence contribute to the acceleration of the low-temperature oxidation route. In addition, at fuel-lean conditions, the excess O_2 is converted to relatively unreactive HO_2 radicals at low temperatures via reaction (R3). Thus, recombination reaction (R3) slows down the reactivity for lean mixtures at low temperatures, which results in higher ignition delay time.

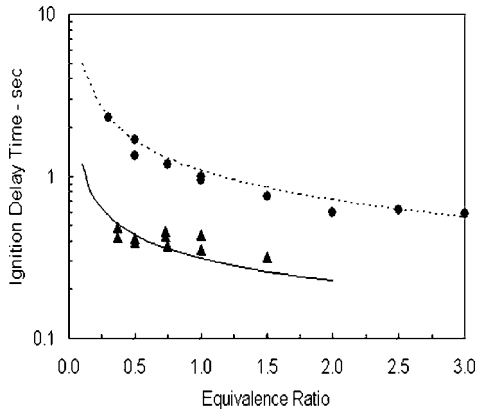


Fig. 11 Kerosene and *n*-heptane ignition delay time as a function of equivalence ratios: (▲) kerosene at 900 K and 17% O_2; (●) *n*-heptane at 865 K and 10% O_2. Solid and dotted lines indicate the modeling results of present study and Curran et al. [25], respectively.

In addition to the kerosene model comparison with the experimental measurements shown in Fig. 11, the predictive capability of the kerosene mechanism is tested with data at different experimental conditions available in the literature. Figure 12 compares the present experimental measurements and model predictions of kerosene/air mixtures against the ignition delay time measurement of Freeman and Lefebvre [33] and Spadaccini and TeVelde [34], who performed experiments in similar tubular flow reactors using Jet A grade kerosene at 1 atm and 20 atm pressures, and 0.5 and 0.7 equivalence ratios, respectively. In Fig. 12, the kerosene/air ignition delay times measured in the present work were obtained at 0.50–0.75 equivalence ratios. The kerosene model predicts the present ignition delay time measurements as well as those of Freeman and Lefebvre [33] very well. The ignition delay time predictions at 20 atm pressure agree well with the experimental measurements of Spadaccini and TeVelde [34] at very low temperatures as shown in Fig. 12. However, it should be noted that the variable length of the premixer used in the experiments [34] appears to have a significant influence on ignition delay time measurements. For example, the ignition delay time measurements shown in Fig. 12 were measured at 20 atm using 116 cm, 70 cm, 24 cm, and 9 cm adjustable mixer/vaporizer length depending on the temperature. The two data points, which deviate far from model predictions (shown in Fig. 12), were obtained at shorter mixing lengths, i.e., 9 cm and 24 cm. Thus, only the experimental data obtained with larger mixing lengths (i.e., 116 cm and 70 cm) are considered for the model comparison, as the shorter mixing lengths seem to introduce significant uncertainty in the ignition delay time measurements.

The Jet-A fuel composition, based on the fuel used in the ex-

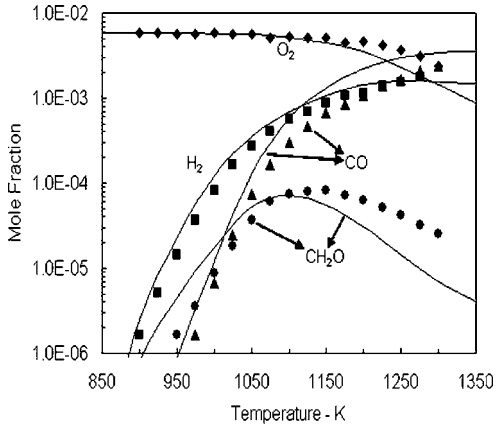


Fig. 13 Model prediction for Jet A-1 reactivity species profile from jet-stirred reactor experiments of Dagaut [14] performed at 1 atm and equivalence ratio of 2.0. The jet-stirred reactor residence time was 70 ms. The symbols and the lines denote the experimental measurements and model predictions, respectively: (●) CH_2O; (▲) CO; (■) H_2; (♦) O_2.

periments [34], is assumed to be 64% *n*-decane, 24% *n*-propylcylohexane, 11% *n*-propylbenzene, and 1% decene for the ignition delay time computation using the kerosene mechanism developed in the present work. Overall, the model predictions agree with experimental measurements fairly well. It is also interesting to note that the ignition delay time measurements of kerosene/air mixtures (0.50–0.75 equivalence ratio) performed at 1 atm pressure in the present study and those of Freeman and Lefebvre [33] collapse into a single exponential function of temperature as noted in Fig. 12. In other words, the ignition delay time measurements of fuel oil No. 1 and Jet-A can be expressed by an Arrhenius relationship with the same "overall" activation energy.

Figures 13 and 14 show the Jet A-1 and JP-8 reactivity profiles, respectively, as a function of reactor temperatures. Dagaut [14] performed jet-stirred reactor experiments at 1 atm varying the temperature of the reactor from 900 K to 1300 K while keeping the residence time of the reactor at 70 ms. The equivalence ratio of the diluted mixture of Jet A-1/O_2/N_2 was maintained at 2.0. Figure 13 compares kerosene model predictions with the experimental measurements of Dagaut [14] assuming the kerosene surrogate composition [14] consists of 74% *n*-decane, 15% *n*-propylbenzene, and 11% *n*-propylcyclohexane for the model simulation. The model agrees with the experimental measurements reasonably well.

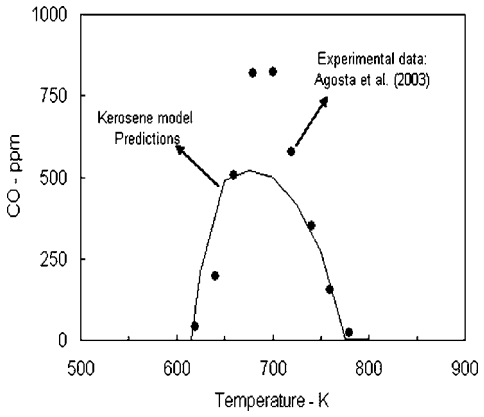


Fig. 14 Model prediction for JP-8 reactivity species profile from the high-pressure plug-flow reactor experiments of Agosta et al. [13] performed at 0.3 equivalence ratio and 12 atm pressure. The flow reactor residence time was 120 ms. The symbols and the lines denote the experimental measurements and model predictions of CO, respectively.

Figure 14 shows the low-temperature reactivity experimental measurement of JP-8 conducted in a high-pressure flow reactor at 12 atm in diluted mixture at an equivalence ratio 0.3 [13]. The chemical composition of the JP-8 used in the experiments was 60% *n*-alkanes, 20% naphthenes, 18% aromatics, and 2% olefins [13]. The model prediction of the surrogate mixture of 60% *n*-decane, 20% *n*-propylcyclohexane, 18% *n*-propylbenzene, and 2% decene reasonably predicts the CO production during low temperature oxidation of JP-8. Therefore, the kerosene mechanism is able to identify the preignition temperature range and the NTC behavior reasonably well. For example, the model predicts the pre-ignition reactivity temperature range between 625 K and 775 K as experimentally observed by Agosta et al. [13]. Also, the model identifies the turnover temperature at which the NTC begins reasonably well with the experimental measurement at approximately 675 K. However, the model needs improvement to predict the maximum CO concentration as observed by the experimental measurements.

Conclusions

The experimental results show that the ignition delay time decreases exponentially as the equivalence ratio of the mixture is decreased at fuel-lean conditions. However, at very high equivalence ratios, the ignition delay time approaches an asymptotic value. Thus, the equivalence ratio fluctuations in the premixer can create a potential scenario for premature ignition of the fuel during premixing due to the presence of pockets of fuel-rich mixture, even though the gas turbines operate under an overall lean condition. Also, the ignition delay time measurement of stoichiometric fuel–oil No. 1/air mixture at 1 atm was comparable with that of kerosene-type Jet A fuel available in the literature.

A detailed chemical kinetics mechanism for kerosene-type fuels is developed using a surrogate mixture of *n*-decane, *n*-propylbenzene, *n*-propylcyclohexane, and decene to represent primary chemical constituents of kerosene, namely, *n*-alkanes, aromatics naphthenes and olefins, respectively. The detailed *n*-decane mechanism that was developed in the present work was validated against a wide range of experimental data from low to high temperatures. The *n*-decane model was able to predict the NTC behavior observed in shock tube ignition delay time measurements fairly well. The surrogate kerosene mechanism predicts the oxidation of Jet A, Jet A-1, JP-8, and fuel oil No. 1 fairly well over a wide range of experimental conditions, including the ignition delay time measurements, species reactivity profile and preignition reactivity profile. Notably, the kerosene model is able to reproduce the NTC behavior observed in the plug-flow reactivity experiments at 12 atm using JP-8 as the fuel. For example, the model correctly predicts the preignition reactivity temperature range between 625 K and 775 K for JP-8 as observed experimentally.

A sensitivity analysis identified the most important reaction steps for *n*-decane oxidation at low, intermediate, and high temperatures. The formation of ketohydroperoxides and their decomposition are the most sensitive reactions during low-temperature oxidation, while reactions involving H_2O_2 become important at intermediate temperatures. At high temperatures, C–C homolysis of *n*-decane and β-scission of decyl radicals become the rate-limiting steps in addition to other common chain carriers.

The addition of a detailed low-temperature submechanism for naphthenes would further improve the model predictions of the kerosene mechanism, as the naphthenes undergo endothermic fuel decomposition at lower temperatures fairly easily. However, the model validation and optimization would be a difficult task, as there are very few low-temperature experimental data available in the literature at the present time.

References

[1] Lefebvre, A. H., 1993, "Combustion Fundamentals," *Gas Turbine Combustion*, Taylor & Francis, Bristol, PA, p. 62.

[2] Lee, J. C. Y., Malte, P. C., and Benjamin, M. A., 2001, "Low NOx Combustion for Liquid Fuels: Atmospheric Pressure Experiments Using a Staged Prevaporizer-Premixer," ASME Paper No. 2001-GT-0081.

[3] Naidja, A., Krishna, C. R., Butcher, T., and Mahajan, D., 2003, "Cool Flame Partial Oxidation and Its Role in Combustion and Reforming of Fuels Cell Systems," Prog. Energy Combust. Sci., **29**, pp. 155–191.

[4] Westbrook, C. K., 2000, "Chemical Kinetics of Hydrocarbon Ignition in Practical Combustion Systems," Proc. Combust. Inst., **28**, pp. 1563–1577.

[5] Bacha, J., Barnes, F., Franklin, M., Gibbs, L., Hemighaus, G., Hogue, N., Lesnini, D., Lind, J., Maybury, J., and Morris, J., 2000, *Aviation Fuels Technical Review*, Chevron Products Company, Chevron U.S.A. Inc., San Ramon, CA.

[6] Edwards, T., and Maurice, L. Q., 2001, "Surrogate Mixtures to Represent Complex Aviation and Rocket Fuels," J. Propul. Power, **17**, pp. 461–466.

[7] Sobel, D. R., and Spadaccini, L. J., 1997, "Hydrocarbon Fuel Cooling Technologies for Advanced Propulsion," J. Eng. Gas Turbines Power, **119**, 344–351.

[8] Huang, H., Spadaccini, L. J., and Sobel, D. R., 2002, "Fuel-Cooled Thermal Management for Advanced Aero Engines," ASME Paper No. GT-2002-30070.

[9] Roubaud, A., Minetti, R., and Sochet, L. R., 2000, "Oxidation and Combustion of Low Alkylbenzenes at High Pressure: Comparative Reactivity and Auto-Ignition," Combust. Flame, **121**, pp. 535–541.

[10] Griffiths, J. F., and Mohamed, C., 1997, "Experimental and Numerical Studies of Oxidation Chemistry and Spontaneous Ignition Phenomena," *Comprehensive Chemical Kinetics*, M. J. Pilling ed., Elsevier, Amsterdam, Vol. 35.

[11] Dagaut, P., Reuillon, M., Boettner, J., and Cathonnet, M., 1994, "Kerosene Combustion at Pressures Up to 40 atm: Experimental Study and Detailed Chemical Kinetic Modeling," Sym. (Int.) Combust., [Proc.], **25**, pp. 919–926.

[12] Pfahl, U., Fieweger, K., and Adomeit, G., 1996, "Self-Ignition of Diesel-Relevant Hydrocarbon-Air Mixtures Under Engine Conditions," Sym. (Int.) Combust., [Proc.], **28**, pp. 781–789.

[13] Agosta, A., Lenhert, D. B., Miller, D. L., and Cernansky, N. P., 2003, "Development and Evaluation of a JP-8 Surrogate that Models Preignition Behavior," *Proceedings of the 3rd Joint Meeting*, Chicago, IL, March 16–19, 2003, The Combustion Institute.

[14] Dagaut, P., 2002, "On the Kinetics of Hydrocarbons Oxidation From Natural Gas to Kerosene and Diesel Fuel," Phys. Chem. Chem. Phys., **4**, p. 2079–2094.

[15] Violi, A., Yan, S., Eddings, E. G., Sarofim, A. F., Granata, S., Faravelli, T., and Ranzi, E., 2002, "Experimental Formulation and Kinetic Model for JP-8 Surrogate Mixtures," Combust. Sci. Technol., **174**, pp. 399–417.

[16] Mawid, M. A., Park, T. W., Sekar, B., and Arana, C. A., 2003, "Development of Detailed Chemical Kinetic Mechanisms for Ignition/oxidation of JP-8/JET-A/JP-7 Fuels," ASME Paper No. GT2003–38932.

[17] Lindstedt, R. P., and Maurice, L. Q., 2000, "Detailed Chemical-Kinetic Model for Aviation Fuels," J. Propul. Power, **16**, pp. 187–195.

[18] Bikas, G., and Peters, N., 2001, "Kinetic Modeling of *n*-Decane Combustion and Autoignition," Combust. Flame, **126**, pp. 1456–1475.

[19] Battin-Leclerc, F., Fournet, R., Glaude, P. A., Warth, V., Come, G. M., and Scacchi, G., 2000, "Modeling of the Gas-Phase Oxidation of *n*-Decane from 550 to 1600 K," Proc. Combust. Inst., **28**, pp. 1597–1605.

[20] Dryer, F. L., 1991, "The Phenomenology of Modeling Combustion Chemistry," *Fossil Fuel Combustion*, W. Bartok and A. F. Sarofim, eds. Wiley, New York.

[21] Zhao, Z., Li, J., Kazakov, A., Dryer, F. L., and Zeppieri, S. P., 2005, "Burning Velocities and a High-Temperature Skeletal Kinetic Model for *n*-Decane," Combust. Sci. Technol., **177**, pp. 89–106.

[22] Zeppieri, S. P., Klotz, S. D., and Dryer, F. L., 2000, "Modeling Concepts for Large Carbon Number Alkanes: A Partially Reduced Skeletal Mechanism for *n*-Decane Oxidation and Pyrolysis," Proc. Combust. Inst., **28**, pp. 1587–1595.

[23] Glaude, P. A., Warth, V., Fournet, R., Battin-Leclerc, F., Scacchi, G., and Come, G. M., 1998, "Modeling of the Oxidation of *n*-Octane and *n*-Decane Using an Automatic Generation of Mechanisms," Int. J. Chem. Kinet., **30**, pp. 949–959.

[24] Lemaire, O., Ribaucour, M., Carlier, M., and Minetti, R., 2001, "The Production of Benzene in the Low-Temperature Oxidation of Cyclohexane, Cyclohexene, and Cyclohexa-1,3-diene," Combust. Flame, **127**, pp. 1971–1980.

[25] Granata, S., Faravelli, T., and Ranzi, E., 2003, "A Wide Range Kinetic Modeling Study of the Pyrolysis and Combustion of Naphthenes," Combust. Flame, **132**, pp. 533–544.

[26] Curran, H. J., Gaffuri, P., Pitz, W. J., and Westbrook, C. K., 1998, "A Comprehensive Modeling Study of *n*-Heptane Oxidation," Combust. Flame, **114**, pp. 149–177.

[27] Benson, S. W., 1968, *Thermochemical Kinetics*, Wiley, New York.

[28] Cohen, N., 1991, "Are Reaction Rate Coefficients Additive? Revised Transition State Theory Calculations for Alkane+OH Reactions," Int. J. Chem. Kinet., **23**, pp. 397–417.

[29] Ribaucour, M., Minetti, R., Sochet, L. R., Curran, J. H., Pitz, W. J., and Westbrook, C. K., 2000, "Ignition of Isomers of Pentane: An Experimental and Kinetic Modeling Study," Proc. Combust. Inst., **28**, pp. 1671–1678.

[30] Gokulakrishnan, P., McLellan, P. J., Lawrence, A. D., and Grandmaison, E. W., 2005, "Kinetic Analysis of NO-Sensitized Methane Oxidation," Chem. Eng. Sci., **60**, pp. 3683–3692.

[31] Lutz, A. E., Kee, R. J., and Miller, J. A., 1990, "Senkin: A FORTRAN Program for Predicting Homogeneous Gas Phase Chemical Kinetics with Sensitivity Analysis," Report No. SAND87-8248, Sandia National Laboratories, Livermore, CA.

[32] Glarborg, P., Kee, R. J., Grcar, J. F., and Miller, J. A., 1986, "PSR: A FORTRAN Program for Modeling Well-Stirred Reactors," Report No. SAND86-8209, Sandia National Laboratories, Livermore, CA.

[33] Freeman, G., and Lefebvre, A. H., 1984, "Spontaneous Ignition Characteristics of Gaseous Hydrocarbon–Air Mixtures," Combust. Flame, **58**, pp. 153–162.

[34] Spadaccini, L. J., and TeVelde, J. A., 1980, "Autoignition Characteristics of Aircraft-Type Fuels," Report No. NASA CR-159886, United Technologies Research Center, East Hartford, CT.

[35] Yetter, R. A., Dryer, F. L., and Rabitz, H., 1991, Combust. Sci. Technol., **79**, pp. 129–140.

[36] Gokulakrishnan, P., Kazakov, A., and Dryer, F. L., 2003, "Comparison of Numerical and Experimental Kinetic Data for Flow Reactor Systems: Mixing Effects," *Proceedings of the 3rd Joint Meeting*, Chicago, IL, March 16–19, 2003, The Combustion Institute.

[37] Horning, D. C., 2001, "A Study of the High-Temperature Autoignition and Thermal Decomposition of Hydrocarbons," Report No. TSD-135, Department of Mechanical Engineering, Stanford University, Stanford, CA.

[38] Ciezki, H., and Adomeit, G., 1993, "Shock-Tube Investigation of Self-Ignition of *n*-heptane–Air Mixtures Under Engine Relevant Conditions," Combust. Flame, **93**, pp. 421–433.

[39] Lifshitz, A., 2001, "Chemical and Combustion Kinetics," *Handbook of Shock Waves*, Vol. 3, Academic, New York.

Laser-Based Investigations of Thermoacoustic Instabilities in a Lean Premixed Gas Turbine Model Combustor

Peter Weigand
e-mail: peter.weigand@dlr.de

Wolfgang Meier

Xuru Duan[1]

Manfred Aigner

Institute of Combustion Technology,
German Aerospace Center (DLR),
Pfaffenwaldring 38-40,
70565 Stuttgart, Germany

Nonintrusive laser-based and optical measurements were performed in a gas turbine model combustor with a lean premixed swirl-stabilized CH_4-air flame at atmospheric pressure. The main objective was to gain spatially and temporally resolved experimental data to enable the validation of numerical CFD results of oscillating flames. The investigated flame was operated at 25 kW and $\phi=0.70$, and exhibited self-excited oscillations of more than 135 dB at ≈300 Hz. The applied measurement techniques were three-dimensional (3D) laser doppler velocimetry (LDV) for velocity measurements, OH^* chemiluminescence yielding information about the heat release and pointwise laser Raman scattering for the determination of joint probability density functions (PDFs) of the major species concentrations, temperature, and mixture fraction. Each of these techniques was applied with phase resolution with respect to the periodic fluctuation of the pressure in the combustion chamber that was measured with a microphone probe. The measurements finally revealed that the mixing of fuel and air in this technical premixing system was strongly affected by the pressure fluctuations leading to changes in equivalence ratio during an oscillation cycle that, in turn, induced the pressure fluctuations. [DOI: 10.1115/1.2718224]

Keywords: gas turbine model combustor, thermoacoustic oscillation, phase-resolved validation measurements

Introduction

Thermoacoustic instabilities in swirling flames are known as a common problem in modern gas turbine combustors operating in lean premixed mode in order to reduce pollutant emissions [1,2]. Numerous studies facing this problem can be found in literature but yielded mostly single-system solutions or only qualitative information about the processes leading to the instabilities, e.g., [3–14]. Despite promising advances, the state of the art in numerical simulation is still not capable of reliably predicting the behavior of oscillating flames. In order to improve our understanding of oscillating flames and to enable the validation of numerical results, spatially and temporally resolved experimental data from combustors with near industrial design on one hand, but with well-defined boundary conditions on the other hand, are urgently needed [15,16]. Measurement techniques using probes (e.g., hot wire, thermocouple, suction probe) are known to interfere with the flow and reaction zones. Additionally, the spatial and temporal resolution is often not sufficient to resolve the fluctuations in turbulent flames. Therefore in this study, nonintrusive optical measurements and, in particular, laser-based techniques were applied in a turbulent swirl flame in order to provide data for the validation and improvement of numerical simulation codes.

For these investigations a premixed swirl-stabilized CH_4-air flame at atmospheric pressure was established in a gas turbine model combustor. The burner nozzle, derived from an industrial design, was a single swirler injector with a conical bluff body on the axis, which was mounted to a square combustion chamber with full optical access. The investigated flame, operated at 25 kW and $\phi=0.70$, exhibited pronounced self-excited oscillations at ~300 Hz. Several complementary measurement techniques were applied separately whose results could be linked together due to the very reproducible behavior of the flame. The pressure fluctuations, which were measured with microphones, were very well repeatable at a nearly constant frequency [17].

Three-dimensional (3D) laser Doppler velocimetry (LDV) was used for the measurement of the flow field, OH^* chemiluminescence to yield information about the location and intensity of the heat release, and pointwise laser Raman scattering for the determination of joint probability density functions (PDFs) of the major species concentrations, temperature, and mixture fraction. In the literature, less costly techniques are reported for measuring variations of the equivalence ratio that are, for example, based on the chemiluminescence emissions from different species [18,19] or on absorption measurements [20,21]. However, laser Raman scattering offers the advantages of high spatial and temporal resolution and the simultaneous detection of additional quantities yielding correlated information of the results. Each technique was applied and evaluated with phase resolution with respect to the periodic fluctuation of the pressure in the combustion chamber. The variations and correlations of the different measured quantities during an oscillation cycle will be shown, revealing the details of the behavior of this flame. The goal of this study was to provide high-quality experimental data of a near technical gas turbinelike burner that exhibited strong pressure oscillations at a dominant frequency for the development and validation of numerical simulation [22,23]. To the knowledge of the authors, such comprehensive data, comprising 3D velocities, temperature, and species concentrations, are not known to date for this kind of flame.

[1]Present address: Southwestern Institute of Physics, P.O. Box 432 Chengdu Sichuan 610041, China

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received June 9, 2006; final manuscript received October 16, 2006. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2006: Land, Sea and Air (GT2006), May 8–11, 2006, Barcelona, Spain. Paper No. GT2006-90300

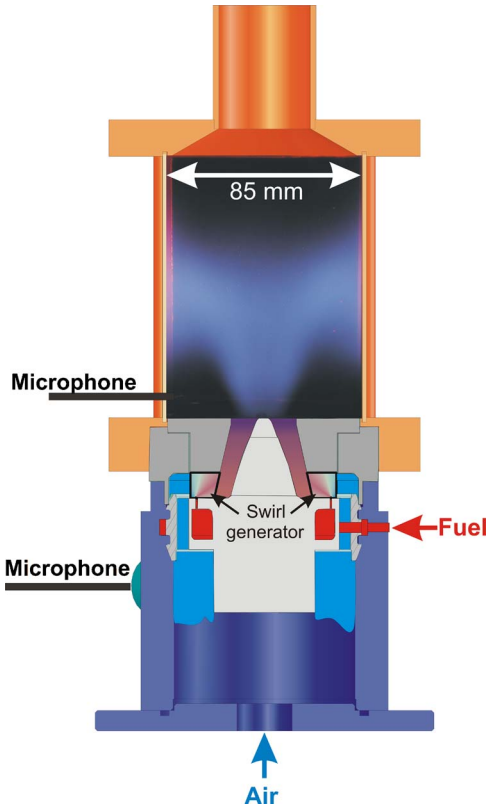


Fig. 1 Schematic of the injector with combustion chamber and photo of the flame

Experimental Setup

For the experiments, a gas turbine model combustor was used that was derived from an industrial design by Turbomeca. In Fig. 1, a schematic of the nozzle design with the combustion chamber is shown. Dry air at ambient temperature is fed via a plenum (80 mm dia) through radial swirler vanes to the burner nozzle. The fuel gas (CH_4) is injected into the airflow through small holes within the radial swirler with high momentum to ensure good mixing before entering the combustion chamber. The burner is operated at 25 kW with an equivalence ratio of 0.70. For these operating conditions, the Reynolds number at the exit of the nozzle, based on the cold flow and the exit diameter (d =29 mm), is ≈35,000 and the swirl number, derived from the velocity measurements at 1.5 mm above the exit, is ≈0.6.

The combustion chamber consists of large quartz windows held by steel posts in the corners, thus creating a confinement with a square section of 85 mm×85 mm and a height of h=114 mm. The exit of the upright combustion chamber is conically shaped, leading to a short central exhaust pipe with a contraction ratio of ≈0.2. The large windows on each side enable unobstructed optical access to nearly the whole flame zone, in particular, close to the nozzle exit. For details of the test rig, see also [22,23].

As can be seen on the photo in Fig. 1, a conically shaped light blue flame was formed that commences already below the exit plane at h=0 mm and hits the walls above h≈30 mm. The flame was constantly flickering over a rather large volume. Despite the squared combustion chamber the visual appearance of the flame was rotationally symmetric. The rotational symmetry of this test rig design has been checked in previous studies with another burner nozzle [24,25]. The time-consuming pointwise laser measurements (LDV and Raman scattering) were thus performed along radial profiles at several heights in one vertical plane. The symmetric profiles resulting from Raman and LDV measurements without phase resolution that were performed from r=−3 mm to 30 mm confirmed the rotational symmetry at least near the axis.

The pressure fluctuations were measured with microphone probes (Brüel & Kjær, 4939 – 1/4 in.), one mounted flush with the inner wall of the plenum and one connected to an air-cooled probe at one of the posts of the combustion chamber at the positions indicated in Fig. 1. The pressure signals from the plenum and from the combustion chamber and their respective power density spectra are shown in Fig. 2. The time signals show the trace of the signal from the plenum being smoother compared to the signal from the combustion chamber. The frequency peaks in both spectra are very pronounced, and no shift of the frequency occurs between the two positions. Therefore, for triggering the measurements, the signal from the plenum was taken as a reference signal because it was less noisy and could thus be used directly without any filtering.

In order to adapt the repetition rate of the intensified ICCD-camera system (≈1 Hz) and the pulsed laser system for Raman (5 Hz) to the oscillation frequency of the flame (f≈300 Hz), a triggering scheme with inhibition times was set up as follows. The negative-to-positive transition of the microphone signal from the plenum triggered a delay generator (SRS, DG535), which itself generated the trigger signal for the laser or the camera systems after a time delay dt1. The inhibition time dt1 was adjusted according to the repetition rate of the laser and camera systems. In order to carry out the measurements at different phase angles, an additional delay time dt2(=0−1/f) was added as indicated in Fig. 3 [25]. The delay time dt2(1/f≈3.3 ms) was small compared to dt1(≈100 ms) and thus permitted stable operation of the laser system.

The measurements were performed at eight phase angles. Although the pressure signal was not perfectly sinusoidal, the assignment of the phase intervals was chosen taking the minimum, maximum, and zero pressure values (and their half intervals) as reference points for the measurements at ph1–ph8. The effective phase triggering and the resulting pressure period compared to a pure sine is shown in Fig. 4.

It can be seen that the pressure rise (maximum at 195 deg) was slightly slower than the pressure drop. Because of a permanently present jitter of the frequency on the scale of approximately ±5 Hz the phase assignment has an uncertainty of maximum ±5 deg at ph8, decreasing with the phase angle. The minimum pressure in the plenum was arbitrarily assigned as ph1=0 deg.

The three velocity components were measured simultaneously by laser doppler velocimetry (LDV) using an Ar^+ laser (Coherent, Innova 90) and two orthogonally positioned Dantec-DISA optics (DISA 55X and DISA flow direction adapter). Commercial camera lenses were used to focus the signals in forward scattering onto three photomultipliers. Signal recording and pretreatment was performed with Dantec BSA-enhanced units (57N20 + 57N35). The seeding particles (TiO_2, d≈0.8 μm) added to the airflow were small enough to follow the large-scale turbulence up to >1 kHz. In order to link the velocity signals to the pressure fluctuations, the BSA units additionally recorded trigger events that were created by the positive zero crossing of the pressure signal of the plenum. In a post-processing step, the velocity signals were assorted by the also documented arrival time of every event into 72 phases (steps of 5 deg) with 10 deg windows (±5 deg). The measurements were performed in one vertical plane along radial profiles at the heights h=1.5 mm, 5 mm, 15 mm, 25 mm, and 35 mm in order to capture the flow field, especially at the nozzle exit. At each position 100,000–200,000 velocity data were recorded. The sorted velocity data were selected according to the phase assignment (ph1–ph8) explained above (Fig. 4). The uncertainty of the velocity measurements for each phase is for the mean value of ≈2.5% and for the rms value of ≈3%.

OH^* chemiluminescence was imaged using an intensified CCD camera (Roper Scientific) equipped with an achromatic ultraviolet

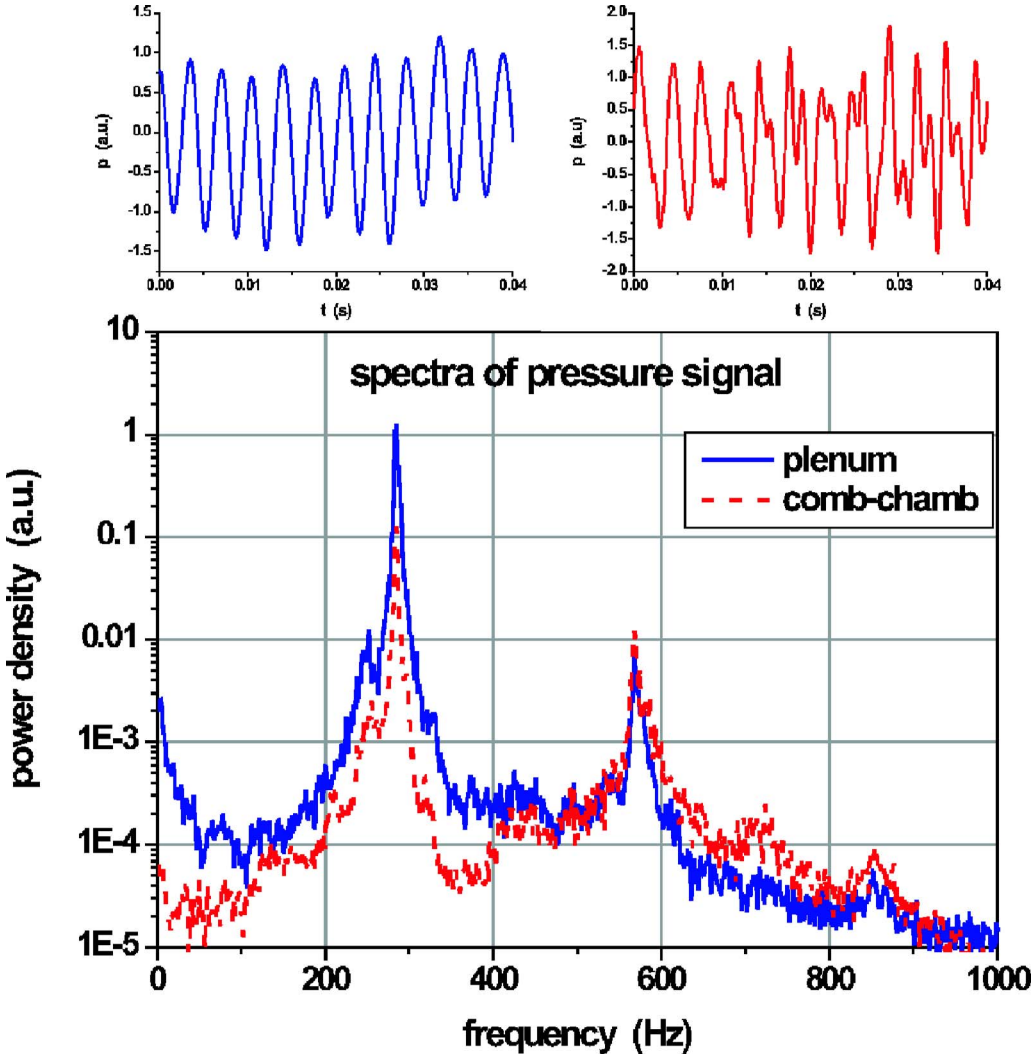


Fig. 2 Time log of the pressure signals and respective spectra from the plenum and the combustion chamber

(UV) lens and an interference filter centered at 312 nm. one-hundred images with an exposure time of 130 μs each were accumulated at each of the eight phase angles. OH^* intensities in lean flames can be regarded as a measure for the heat release rate [26–28]. Thus from these images, information can be deduced

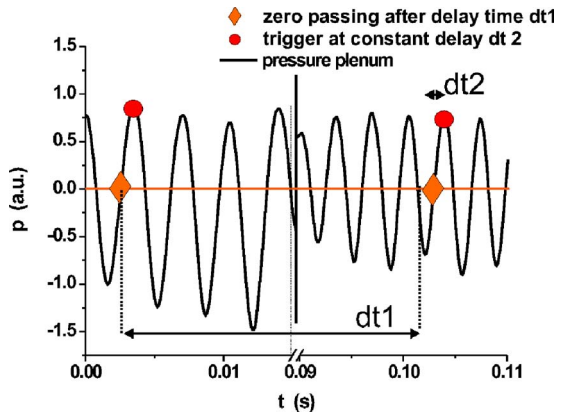


Fig. 3 Pressure signal of the plenum with trigger scheme for the pulsed measurements

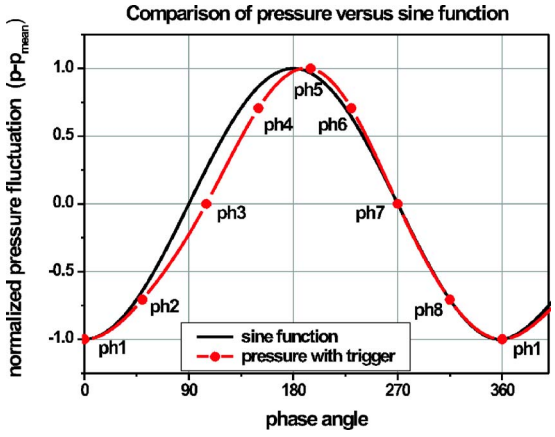


Fig. 4 Pressure oscillation compared to a pure sine function; the dark (black) curve shows a sine function and the light (red) curve displays the pressure signal of the plenum. The markers ph1–ph8 indicate the assigned phase angles at which measurements were performed. These were ph1=0 deg, ph2=50 deg, ph3=105 deg, ph4=150 deg, ph5=195 deg, ph6=230 deg, ph7=270 deg, and ph8=315 deg.

about the phase-resolved variations in heat release intensity and the location and extension of the flame zone. Even though this technique is integrating line of sight, spatially resolved information can be gained by deconvolution, taking advantage of the rotational symmetry of the flame. This results in a "quasi two-dimensional (2D) image" of the center plane where the local distribution of the intensity of the heat release can be identified more clearly than in the integral view.

Laser Raman scattering was performed with a flash lamp pumped dye laser (Candela LDFL 20, 489 nm, pulse energy 3 J, pulse duration 3 μs, repetition rate 5 Hz), and detecting the signals with a high-resolution spectrograph (SPEX 1802, $f=1$ m, dispersion ≈ 0.5 nm/mm) and one photomultiplier for each species in the detection plane. In order to correct for laser-induced emissions possibly arising from aromatics or C_2 in fuel-rich regions, the background signal was detected with additional photomultipliers. Because the background signal in this flame was very low at all measurement positions, no correction had to be applied. The detected probe volume was ≈ 0.6 mm dia with a length of 0.6 mm. From the signals, the species number densities for all major species of the CH_4-air flame (O_2, N_2, CH_4, H_2, CO, CO_2, H_2O) were derived using calibration measurements. From there, the mixture fraction could be calculated and the temperature was deduced from the total number density and the knowledge of the pressure via the ideal gas law [29,30]. Because of the optical setup, the closest possible distance between the burner exit and the probe volume was $h=6$ mm. Radial profiles were measured at $h=6$ mm, 15 mm, 25 mm, 35 mm, and 60 mm, focusing on the region near the nozzle exit. At each of the ≈ 50 measurement positions, 400 samples were taken for each of the eight phase angles yielding the joint probability density functions (PDFs), from which the mean values and rms fluctuations were derived. The mean values of temperature and mixture fraction obtained are within an uncertainty of $<4\%$ [24].

Results and Discussion

The comparison of the simultaneously measured pressure signals from the plenum and the combustion chamber (not shown) revealed that at 290–300 Hz, the pressure oscillation in the combustion chamber runs ahead of the pressure in the plenum by ≈ 80 deg. That means for ph3 the pressure was nearly at its maximum in the combustion chamber. This should be kept in mind when interpreting the results presented hereafter.

Figure 2 shows a typical example of the frequency spectra that were randomly taken during the laser measurements. The peak frequency here is at $f=290$ Hz and more than three orders of magnitude higher than the broadband noise level. From Fig. 2, it can also be seen that in the plenum the ratio of the main peak to the second harmonic is a factor of nearly ten higher, compared to the ratio measured in the combustion chamber and the contribution of higher harmonics is generally negligible. The sound pressure level for the peak frequency in the combustion chamber was evaluated to be about 135–138 dB.

In order to give an overall impression of the flow field, the combined mean axial (u) and radial (v) velocities are shown in Fig. 5 as a vector plot, with the zones of negative and positive u velocities marked dark red and light blue, respectively. The color assignment was created by interpolating the mean axial velocities.

The negative u velocities in the center clearly identify the inner recirculation zone (IRZ). The IRZ reveals surprisingly high negative velocities, and the extension in axial direction reaches far higher than the maximum measured height of 35 mm, so that unfortunately the upper limit of the IRZ is not captured. The existence of an outer recirculation zone (ORZ) can be deduced indirectly from the inward directed v velocities at $h=1.5$–15 mm and also from the color of the interpolation becoming reddish toward the borders. The ORZ seems to be thin and to stick very close to the walls; so that it could not clearly be detected by

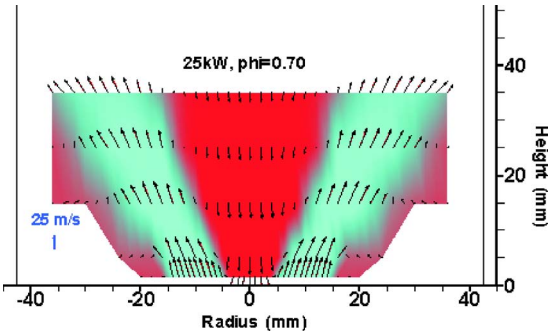


Fig. 5 Vector plot of the mean uv velocities; the half profiles were mirrored to yield a more complete impression of the flow; zones with negative and positive u velocities are marked dark red and light blue, respectively. The coloring is based on interpolation of the measured mean u velocities (using the program Tecplot 10).

negative axial velocities because the LDV measurements could not be performed closer than 5–6 mm to the wall.

The transport of hot gases and combustion products back to the flame root by the IRZ is usually the main mechanism that is responsible for the stabilization of this kind of swirl flames. If the IRZ undergoes major variations, the flame may thus be prone to unstable operation. In Fig. 6, the variation of the extension of the IRZ during a pressure cycle is shown by plotting the contour of $u=0$ m/s for different phases. As can be seen, the form and location of the IRZ is changing drastically during the pressure cycle in the range of the measured heights up to 35 mm.

At ph8, the IRZ is lifted by ≈ 5 mm above the exit plane corresponding to the maximum inlet velocities in the region of $r=5$–15 mm (displayed in Fig. 7). Subsequently, the IRZ is moving upstream again, and at ph2 it is already reaching below the lowest measuring height of $h=1.5$ mm. The radial extension at $h=1.5$ mm is now continuously increasing until ph6, and although it could not be measured, it seems plausible to assume that the IRZ attaches to the bluff body and increasingly extends upstream for an unknown distance into the nozzle. It was also observed visually that the luminous flame zone already starts below the nozzle exit. The pumping motion of the IRZ is related to the changes of the inflow velocities but also to the changes of the ratio of circumferential and axial velocities thus modifying the swirl strength.

In Fig. 7, the radial profiles at $h=5$ mm for the axial and circumferential velocities for the same phases as in Fig. 6 are shown.

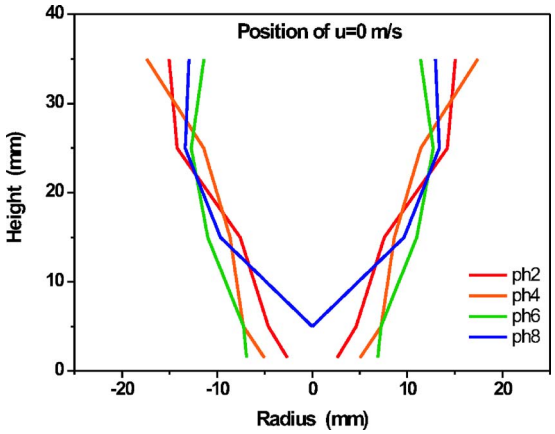


Fig. 6 Contours of $u=0$ m/s for different phase angles indicating the changes of the inner recirculation zone during a pressure cycle (values mirrored for better visualization)

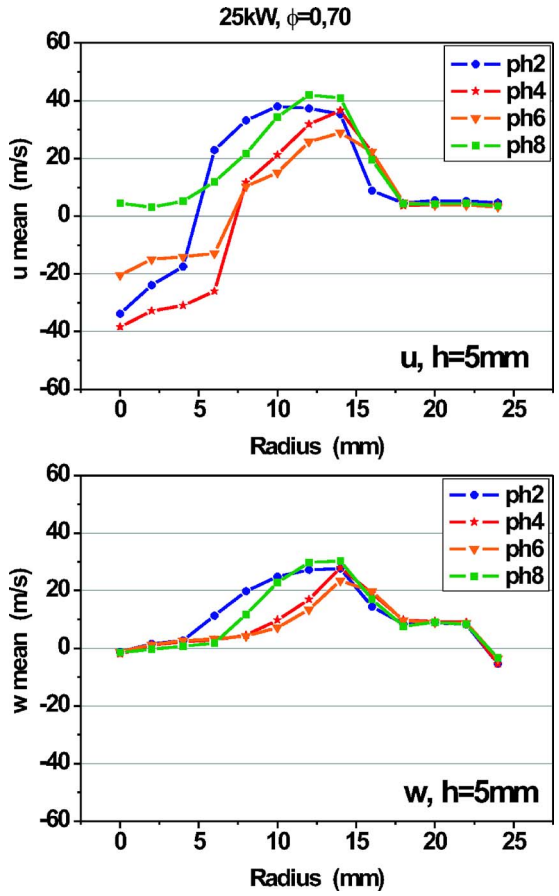


Fig. 7 Radial profiles of axial velocity (top) and circumferential velocity (bottom) for different phase angles at $h=5$ mm

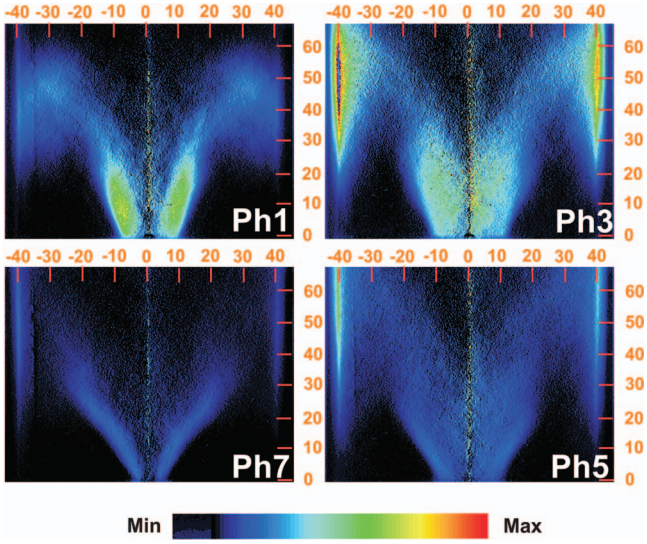


Fig. 8 Deconvoluted OH* chemiluminescence images (averaged over 200 images) for different phases; the sharp gradients and the extremely high intensities at the walls are an artifact of the deconvolution routine, but there are, in reality, high intensities near the walls at ph3 (in particular from ph2 to ph4 and still at ph5)

The phase-resolved analysis of the velocities reveals strong variations of the axial velocity on the axis, varying between −40 m/s and +3 m/s at $h=5$ mm during one cycle. Multiplying the velocities with the corresponding centered ring section and summing up yields the integrated volume flow for the different phases. At ph8, the integrated inflow for $h=5$ mm (integrated up to $r=24$ mm) is at its maximum, pushing the IRZ out and away from the nozzle exit. This correlates reasonably with the minimum pressure in the combustion chamber at ph7. The circumferential velocity at ph8 is increased compared to the preceding phase, thus enforcing the vortex breakdown leading to the IRZ. With increasing pressure in the combustion chamber (ph8–ph3), the axial velocity is reduced while the integrated circumferential velocity reaches its maximum at ph3.

As a result the ratio of circumferential and axial velocity, and thus the swirl number, increases. This enhances the reverse flow, and the IRZ reattaches to the bluff body in the nozzle exit. At ph2, the IRZ is already well established with the negative axial velocity on the axis coming close to maximum. The IRZ moving down into the nozzle displaces the inflow region outward. At about ph5, the integrated inflow is at its minimum. This is a phase delay of 90 deg with respect to the pressure minimum in the combustion chamber at ph3. The circumferential velocity is reduced from ph4 onward, while the axial velocity rises again from ph6 on, leading finally to a decreased backflow in the IRZ from ph6 to ph8. At ph8 finally, the inflow pushes the IRZ away from the nozzle exit, restarting the cycle.

The effect of this pumping motion on the combustion can be identified by imaging the chemiluminescence of OH* radicals. This measuring technique has frequently been applied to characterize periodic oscillations of the heat release [27,31,32]. In order to yield more spatially resolved information, the images were deconvoluted using the three-point interpolation Abel inversion method (for details, see [33]) taking advantage of the rotational symmetry of the flame. As a result, a quasi-2D cut through the flame is obtained. In Fig. 8, the deconvoluted images of OH* for the phase angles ph1, ph3, ph5, and ph7 are displayed. First, from these images it can be seen that for all phases combustion has already started at the nozzle exit. Second, it becomes evident that the heat release intensity and the location of the flame zone vary drastically during one cycle. Looking at the intensity distribution, it has to be mentioned that the sharp gradients and the extremely high intensities at the walls are an artifact of the deconvolution routine. Because of the sharp cutoff at the wall, the gradient is overestimated by the program. However, there are high intensities near the walls from ph2 to ph4, when the main flame zone is at $h\approx 35-60$ mm and $r>20$ mm and the flame is burning against the walls. At ph5, this is still the case but already very alleviated

Starting at ph7 (≈minimum pressure in combustion chamber), the OH* chemiluminescence is weak, indicating that combustion intensity is low. Significant OH* chemiluminescence, and thus reaction, occurs only in the center near the shear layer of the IRZ and inflow. From there on, OH* increases rapidly, indicating that the reaction rate is strongly enhancing. At ph1, the highest intensities—representing the main flame zone—are near the nozzle exit, where the fresh gases, coming with the high inflow at ph8, meet the hot gases of the IRZ and ignite. This flame zone moves downstream and finally reaches the walls at ph2–ph3. At ph3, the flame is burning very intensely near the walls but also in the core of the IRZ, resulting in the maximum integrated heat release. Because of the strong heat release, the maximum pressure in the combustion chamber is also reached at ph3. After ph3, the OH* intensity is decreasing and the flame degrades until at ph7 the minimum is reached. The pressure is following this mitigation. The remaining flame zone in the shear layer near the IRZ at ph7 acts as an ignition source for the fresh gas coming next with ph8, thus restarting the cycle.

In Fig. 9, the phase-dependent normalized OH* signal integrated over the detected image is plotted together with the normalized pressure fluctuation in the combustion chamber. As can be seen, the heat release and pressure fluctuations are very well in phase, thus fulfilling the Rayleigh criterion for self sustained os-

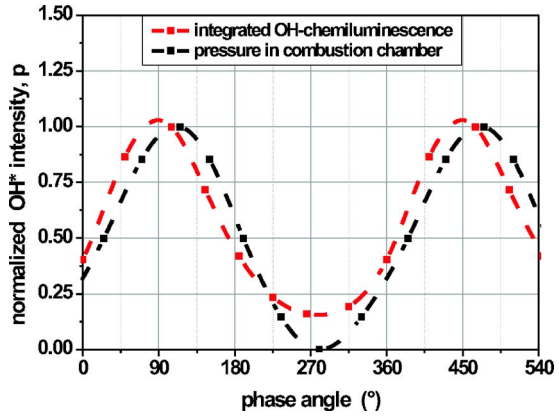


Fig. 9 Normalized integrated OH* chemiluminescence and normalized pressure fluctuation in the combustion chamber

cillations in flames. The integrated intensity of the OH^* chemiluminescence varies between 100% and 15%, demonstrating the strong fluctuation of the heat release during a pressure cycle.

The Raman results with the joined information of temperature and species concentrations help to understand this behavior. In Fig. 10, the 2D temperature distribution for different phase angles is shown. These images are obtained by interpolating single point measurements at radial positions, indicated by a star at phase 1, ranging between $r=0-30$ mm at the heights $h=6$ mm, 15 mm, 25 mm, 35 mm, and 60 mm.

Although the step between $h=35$ and $h=60$ is quite coarse, the general information can be regarded as representative of the process. At ph7 and ph1, one identifies the cold gases penetrating deeply into the combustion chamber (<1000 K up to 20 mm and ≈ 28 mm, respectively). This corresponds to the high inflow velocities around ph8 described in Fig. 7. From ph1 to ph5, the temperature in the area $r=10-30$ and $h=10-40$ is increasing according to the combustion represented by the OH^* intensities shown in Figs. 8 and 9. The temperature lags the heat release (OH^* signal) by about 60–90 deg. While the OH^* signal reaches maximum at ph3, the temperature maximum is reached at ph5. As the heat release decreases from ph4 to ph7, the temperature in the described zone follows with the already mentioned delay. At the same time, the temperature in the IRZ is increasing, what may be due to a still ongoing reaction of the gases that were previously transported into the IRZ. These recirculated hot gases finally mix with the fresh gases in the shear layer of IRZ and incoming gas flow and ignite them. The strong increase in combustion intensity at the nozzle exit at ph1, which can be seen in Fig. 8, results in the temperature rise near the shear layer ≈ 90 deg later at ph3. The temperature distribution also reveals some information about the variation of the extension of the IRZ. As the IRZ mainly consists of hot reacted gases, the changing boundary of the hot gases in the core of the flame gives a rough idea of the variation of the extension of the IRZ.

The concentration measurements help one to understand why this periodic fluctuations of the temperature and heat release show up in a premixed flame. In Fig. 11, the 2D distribution of the concentration of the fuel (CH_4) is shown together with the vectors of the combined axial and radial velocities for different phase angles. It is found that the equivalence ratio is considerably varying. At ph7 the concentration of fuel in the inflow of fresh gas near the nozzle exit (corresponding to the low temperature region in Fig. 10) is high. Surprisingly, the fresh gas mixture here contains more fuel than corresponding to the equivalence ratio of the premixed gas mixture ($\phi=0.70$, 7.36% CH_4) and even exceeds the stoichiometric value of 10.5% CH_4. It is even more surprising that this high concentration of fuel can still be detected at rather large heights (up to 15 mm at ph8, not shown) because the mixing of

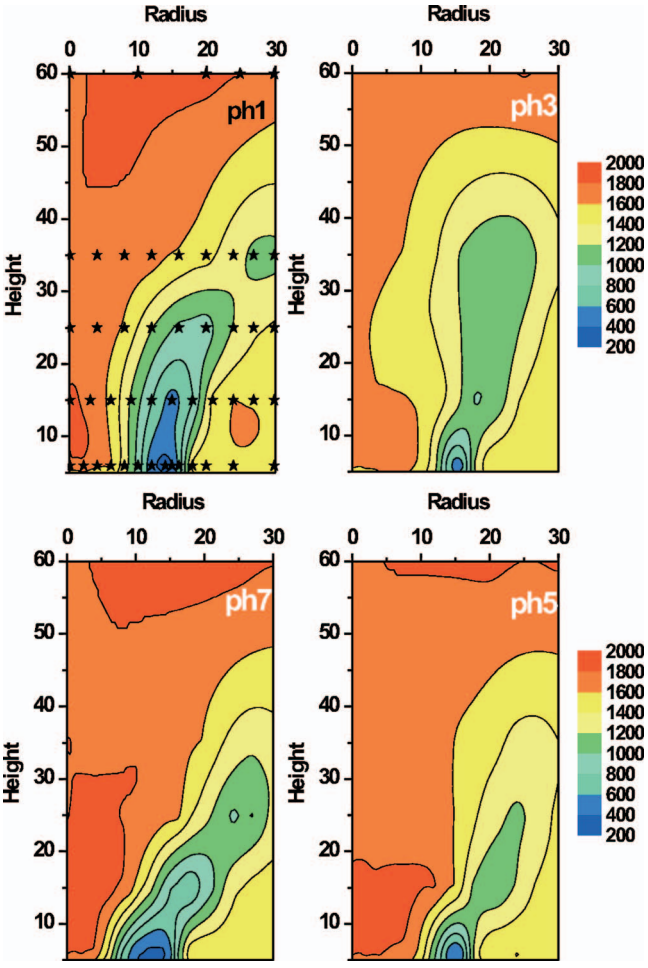


Fig. 10 2D temperature distribution from interpolation (using the program ORIGIN 7.5) of single-point Raman measurements (averaged over 400 sample at each position) at the heights *h* =6 mm, 15 mm, 25 mm, 35 mm, and 60 mm for different phase angles. The measurement positions are indicated by stars at ph1.

fuel and air was accomplished quite a distance upstream in the swirler passages at high velocities of fuel and air.

As the CH_4 distribution and the OH^* images in Fig. 8 show, this fuel-rich gas mixture does not ignite immediately but is transported downstream and diluted before, at heights $h>30$ mm, the main flame zone is established at about ph2. From ph1 to ph5, the fuel in the upper regions ($h>15$ mm) vanishes in accordance with the increase of the reaction rate as was discussed above. At ph3, only little fuel is fed with the incoming flow and the concentration of fuel in the inflow is now below the expected mean for a perfectly mixed fuel-air mixture. At ph5, the remaining concentration of CH_4 at $h>15$ mm is low and, consequently, the heat release is diminishing as discussed at Fig. 8. At ph5, one can already identify the fuel concentration in the incoming flow becoming higher and, at ph7 and ph8, the concentration again exceeds the stoichiometric ratio closing the cycle.

The variation in fuel concentration of the incoming flow leads to the conclusion that the mechanism responsible for the oscillation is not purely acoustically triggered, but also due to the fluctuating fuel-air mixture. As the pressure drop of the airflow through the swirler ($\Delta p \approx 0.1$ kPa) is far beyond the pressure drop of the injected fuel ($\Delta p \approx 2.2$ kPa), it is assumed that the airflow response to the pressure fluctuation in the combustion chamber is more sensitive than the fuel flow. To check this assumption, an estimation of a characteristic time of the flow was performed,

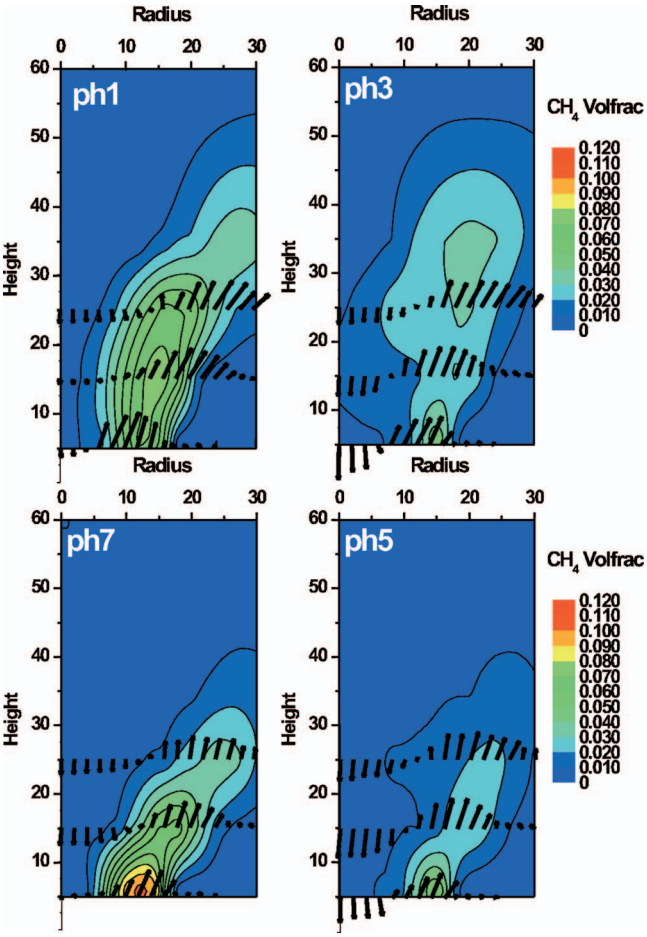


Fig. 11 2D distribution of CH_4 volume fraction interpolated from single-point Raman measurements (using the program ORIGIN 7.5), combined with the *uv* vectors of the LDA measurements for different phase angles

going from the mixing position in the swirler to the nozzle exit, where the variation in fuel concentration could be measured at a height of 6 mm. For this distance, the fluid needs approximately half the time of a pressure cycle. This leads to the following possible explanation: At the time ph3, when the pressure in the combustion chamber is high, the inflow velocity is low. The reduced inflow is mainly caused by a reduced airflow while the fuel flow into the airflow remains nearly constant. Because of this effect, a fuel-rich mixture is created at the mixing point in the swirler passage. This fuel-rich mixture is then transported with the now accelerated flow toward the nozzle exit, where it appears a half-period later at ph7–ph8. During this time, the fuel-air mixture is becoming leaner at the mixing point because now more air than the average is going through the swirler, showing up at the nozzle exit at about ph3–ph4. The fuel-rich gas mixture entering the combustion chamber at ph8 undergoes the cycle of combustion oscillation as described above: ignition at ph1 and maximum heat release at ph3 leading to the increase in pressure with the maximum pressure at ph3. This, in turn, causes the drop in the inflow, which again is the reason for the creation of a fuel rich mixture.

Summary and Conclusion

Nonintrusive optical measurements were performed at an oscillating premixed swirl flame at atmospheric pressure in order to characterize the flame during the cyclic variation. The measurements were coupled to the pressure oscillations recorded by microphones. The 3D flow field was measured with LDV, and large variations during a cycle were identified. In order to gain information about the heat release, the OH^* chemiluminescence of the flame was measured by an ICCD camera. The variations of the heat release showed good accordance with the pressure fluctuations as is to be expected by the Rayleigh criterion. The most interesting results were detected by laser Raman spectroscopy. The results revealed that the pumping behavior of the flame is mainly due to changes in equivalence ratio of the fuel-air mixture entering the combustion chamber and not primarily due to acoustical effects. The cycle leading to the pressure fluctuations can be described as follows, starting at low pressure in the combustion chamber (ph7): The relatively low pressure in the combustion chamber allows higher inflow velocities of the fresh gas. The fresh gas mixture does not ignite directly due to the high velocity and the ignition delay at that temperature. The mixing of fresh gas and hot gas from the inner recirculation zone (IRZ) taking place at the shear layer is weak because the IRZ at that moment is pushed upward into the combustion chamber. The combustion strongly increases when the gas mixture is already near the walls (ph2–ph3), where the velocities are reduced and enough time has elapsed for ignition. The heat release taking place now produces a fast pressure rise, and the elevated pressure, in turn, reduces the inflow of the fresh gas mixture. As the pressure drop for the fuel at the injection whole is much higher than the pressure drop for the air in the swirler passages, the response of the two flows is different. The fuel flow presumably remains more or less constant, while the airflow is accelerated and retarded with the pressure oscillation in the combustion chamber. This leads to the effect that the fuel-air ratio that is created in the swirler is varying. The fuel-rich mixture that is created in the swirler during the high-pressure phase at ph2–ph4 is transported through the nozzle and enters the combustion chamber at ph7–ph1. Due to the above-mentioned reasons, the heat release is delayed coming to the maximum at about ph3 creating again the pressure rise. As a summary, one can state that the pressure fluctuation is mainly caused by the variation of the fuel-air mixture, which itself is caused by the pressure fluctuation. This mechanism has been reported in the literature before (see, for example, [7,34]); however, the experimental results presented here yield a quantitative and very detailed description of it and therefore allow a deeper insight into the dynamic behavior of the flame. It also leads to the conclusion that a high-pressure drop of the fuel supply is not sufficient to guarantee a homogenous mixture of fuel and air.

Because of its unique quality and comprehensiveness, this data set is optimally suited for the validation and improvement of numerical codes for technically premixed swirling flames. The data may be available together with the corresponding geometry details of the burner on request (corresponding author).

Acknowledgment

The financial support of this work by the EU project PRECINSTA Grant No. ENK5-CT-2000-00060 is gratefully acknowledged by the authors. We also like to thank Turbomeca SA for the provision of the burner nozzle in the frame of this project. Furthermore, we thank Bernhard Lehmann at DLR Berlin for executing the LDA measurements and Berthold Noll for stimulating discussions.

References

[1] Candel, S., 2002, "Combustion Dynamics and Control: Progress and Challenges," Proc. Combust. Inst., **29**, pp. 1–28.
[2] McManus, K. R., Poinsot, T., and Candel, S. M., 1993, "A Review of Active Control of Combustion Instabilities," Prog. Energy Combust. Sci., **19**, pp. 1–29.
[3] Polifke, W., Fischer, A., and Sattelmayer, T., 2003, "Instability of a Premix Burner With Non-monotonic Pressure Drop Characteristic," ASME J. Eng. Gas Turbines Power, **125**, pp. 20–27.
[4] Paschereit, C. O., Gutmark, E., and Weisenstein, W., 1998, "Structure and Control of Thermoacoustic Instabilities in a Gas-Turbine Combustor," Combust. Sci. Technol., **138**, pp. 213–232.
[5] Putnam, A. A., 1971, *Combustion-Driven Oscillations in Industry*, Elsevier, New York.

[6] Di Benedetto, A., Marra, F. S., and Russo, G., 2002, "Spontaneous Oscillations in Lean Premixed Combustion," Combust. Sci. Technol., **174**, pp. 1–18.
[7] Lieuwen, T., Torres, H., Johnson, C., and Zinn, B. T., 2001, "A Mechanism of Combustion Instabilities in Lean Premixed Gas Turbine Combustors," ASME J. Eng. Gas Turbines Power, **123**, pp. 182–189.
[8] Dawson, S., and Fitzpatrick, J. A., 2000, "Measurement and Analysis of Thermoacoustic Oscillations in a Simple Dump Combustor," J. Sound Vib., **230**(3), pp. 649–660.
[9] Umurhan, O. M., 1999, "Exploration of Fundamental Matters of Acoustic Instabilities in Combustion Chambers," Center for Turbulent Research, Annual Briefs, pp. 85–98.
[10] Emiris, I., and Whitelaw, J. H., 2003, "Control of Combustion Oscillations," Combust. Sci. Technol., **175**, pp. 157–184.
[11] Lepers, J., Krebs, W., Prade, B., Flohr, P., Pollardo, G., and Ferrante, A., 2005, "Investigation of Thermoacoustic Stability Limits of an Annular Gas Turbine Combustor Test-Rig With and Without Helmholtz Resonators," ASME Paper No. GT2005-68246.
[12] Lieuwen, T., and Zinn, B. T., 1998, "The Role of Equivalence Ratio Oscillations in Driving Combustion Instabilities in Low NO_x Gas Turbines," Sym. (Int.) Combust., [Proc.], **27**, pp. 1809–1816.
[13] Keller, J. O., Barr, P. K., and Gemmen, R. S., 1994, "Premixed Combustion in a Periodic Flow Field. Part I: Experimental Investigation," Combust. Flame, **99**, pp. 29–42.
[14] Barr, P. K., and Keller, J. O., 1994, "Premixed Combustion in a Periodic Flow Field. Part II: The Importance of Flame Extinction by Fluid Dynamic Strain," Combust. Flame, **99**, pp. 43–52.
[15] Vanoverberghe, K. P., and van den Bulck, E. V., 2003, "Confined Annular Swirling Jet Combustion," Combust. Sci. Technol., **175**, pp. 545–578.
[16] Ji, J., and Gore, J. P., 2002, "Flow Structure in Lean Premixed Swirling Combustion," Proc. Combust. Inst., **29**, pp. 861–867.
[17] Weigand, P., Duan, X. R., Meier, W., Meier, U., Aigner, M., and Bérat, C., 2005, "Experimental Investigations of an Oscillating Lean Premixed CH_4-air Swirl Flame in a Gas Turbine Model Combustor," *Proc. of the European Combustion Meeting 2005*, Louvain-La-Neuve, Belgium, The Combustion Institute, p. 235.
[18] Hardalupas, Y., and Orain, M., 2004, "Local Measurements of the Time-Dependent Heat Release Rate and Equivalence Ratio Using Chemiluminescent Emission From a Flame," Combust. Flame, **139**, pp. 188–207.
[19] Ikeda, Y., Kojima, J., Nakajima, T., Akamtsu, F., and Katsuki, M., 2000, Proc. Combust. Inst., **28**, pp. 343–350.
[20] Nguyen, Q.-V., 2002, "Measurement of Equivalence Ratio Fluctuations in a Lean Premixed Prevaporized (LPP) Combustor and its Correlation to Combustion Instability," ASME Paper No. GT2002-30060.
[21] Mongia, R., Dibble, R., and Lovett, J., 1998, "Measurement of Air Fuel Ratio Fluctuations Caused by Combustor Driven Oscillations," ASME Paper No. 98-GT304.
[22] Lartigue, G., Meier, U., and Bérat, C., 2004, "Experimental and Numerical Investigation of Self-excited Combustion Oscillations in a Scaled Gas Turbine Combustor," Appl. Therm. Eng., **24**, pp. 1583–1592.
[23] Roux, S., Lartigue, G., Poinsot, T., Meier, U., and Bérat, C., 2004, "Studies of Mean and Unsteady Flow in a Swirled Combustor Using Experiments, Acoustic Analysis and Large Eddy Simulation," Combust. Flame, **141**, pp. 40–54.
[24] Weigand, P., Meier, W., Duan, X. R., Stricker, W., and Aigner, M., 2006, "Investigations of Swirl Flames in a Gas Turbine Model Combustor. Part I: Flow Field, Structures, Temperature and Species Distribution," Combust. Flame, **144**, pp. 205–224.
[25] Duan, X. R., Meier, W., Weigand, P., and Lehmann, B., 2005, "Phase-Resolved Laser Raman Scattering and Laser Doppler Velocimetry Applied to Periodic Instabilities in a Gas Turbine Model Combustor," Appl. Phys. B: Lasers Opt., **80**, pp. 389–396.
[26] Haber, L. C., Vandsburger, U., Saunders, W. R., and Khanna, V. K., 2000, "An Examination of the Relationship Between Chemiluminescent Light Emissions and Heat Release Rate Under Non-Adiabatic Conditions," *Proc. of Int. Gas Turbine Institute*, Munich, ASME, New York, Paper No. 0121.
[27] Lee, J. G., and Santavicca, D. A., 2003, "Experimental Diagnostics for the Study of Combustion Instabilities in Lean Premixed Combustors," J. Propul. Power, **19**, pp. 735–750.
[28] Docquier, N., and Candel, S., 2002, "Combustion Control and Sensors: A Review," Prog. Energy Combust. Sci., **28**, pp. 107–150.
[29] Keck, O., Meier, W., Stricker, W., and Aigner, M., 2002, "Establishment of a Confined Swirling Natural Gas—Air Flame as a Standard Flame: Temperature and Species Distributions From Laser Raman Measurements," Combust. Sci. Technol., **174**, pp. 73–107.
[30] Bergmann, V., Meier, W., Wolff, D., and Stricker, W., 1998, "Application of Spontaneous Raman and Rayleigh Scattering and 2-D LIF for the Characterization of a Turbulent CH4/H2/N2 Jet Diffusion Flame," Appl. Phys. B: Lasers Opt., **66**, pp. 489–502.
[31] Lee, S.-Y., Seo, S., Broda, J.-C., Pal, S., and Santoro, R. S., 2000, "An Experimental Estimation of Mean Reaction Rate and Flame Structure During Combustion Instability in a Lean Premixed Gas Turbine Combustor," Proc. Combust. Inst. **28**, pp. 775–782.
[32] Giezendanner, R., Keck, O., Weigand, P., Meier, W., Meier, U., Stricker, W., and Aigner, M., 2003, "Periodic Combustion Instabilites in a Swirl Burner Studied by Phase-Locked Planar Laser-Induced Fluorescence," Combust. Sci. Technol., **175**, pp. 721–741.
[33] Dasch, J. C., 1992, "One-Dimensional Tomography: A Comparison of Abel, Onion Peeling, and Filtered Backprojection Methods," Appl. Opt., **31**(8), pp. 1146–1152.
[34] Auer, M. P., Hirsch, C., and Sattelmayer, T., 2001, "Influence of the Interaction of Equivalence Ratio and Mass Flow Fluctuations on Flame Dynamics," ASME Paper No. GT2005-68373.

A Novel Approach to Predicting NO_x Emissions From Dry Low Emissions Gas Turbines

Khawar J Syed
Siemens Industrial Turbomachinery Ltd.,
P. O. Box 1,
Lincoln LN5 7FD, UK

Kirsten Roden
Siemens AG
Guenther-Scharowsky-Str. 1,
D-91058 Erlangen, Germany

Peter Martin
Siemens Industrial Turbomachinery Ltd.,
P. O. Box 1,
Lincoln LN5 7FD, UK

An empirical modeling concept for the prediction of NO_x emissions from dry low emissions (DLE) gas turbines is presented. The approach is more suited to low emissions operation than are traditional approaches (Lefebrre, A. H., 1998, Taylor and Francis, New York). The latter, though addressing key operating parameters, such as temperature and pressure drop, do not address issues such as variation in fuel/air distribution through the use of multifuel stream systems, which are commonly applied in DLE combustors to enable flame stability over the full operating range. Additionally, the pressure drop dependence of NO_x in such systems is complex, and the exponent of a simple pressure drop term can vary substantially. The present approach derives the NO_x model from the equations that govern the NO_x chemistry, the fuel/air distribution and the dependence of the main reaction zone on its controlling parameters. The approach is evaluated through comparing its characteristics to data obtained from high-pressure testing of a DLE combustor fueled with natural gas. The data were acquired at a constant pressure and preheat temperature (14 Bara and 400°C) and a range of flame temperatures and flow rates. Though the model is configured to address both relatively fast and slow NO_x formation routes, the present validation is conducted under conditions where the latter is negligible. The model is seen to reproduce key features apparent in the data, in particular, the variable pressure drop dependence without any ad hoc manipulation of a pressure drop exponent. [DOI: 10.1115/1.2718564]

1 Introduction

There is a need for emissions prediction algorithms that are both accurate and suitable for incorporation into, for example, combustor design codes and on-line emissions prediction tools. To be effective, such algorithms must be accurate, as well as being applicable over a wide range of operating and ambient conditions.

In addition to the above requirements, additional consideration has to be given to the following:

a. The amount of data that can feasibly be acquired to establish an empirically based tool will be relatively small; it will not be possible to measure emissions on machines operating over the complete operating envelope, over the complete set of ambient conditions, and over all build variants.
b. The model cannot rely totally on fundamental theory, since the complex physical and chemical processes that underlie emissions formation and consumption cannot be represented satisfactorily with a simple tool.
c. There will be uncertainty in experimental measurements, particularly from machines where combustor operating conditions have to be calculated rather than measured.

Given the above requirements and considerations, it is clear that an algorithm based on simple mathematical curve fitting, or even neural networks, will suffer from insufficient data. At the other extreme, a model based on the fundamental processes will be too cumbersome for the requirements. Therefore, within the present work an intermediate route is adopted, where the fundamental processes are analyzed to identify the key processes. The representation of these processes is then simplified, based on a clear set of assumptions. The result is a tractable model, which requires empiricism to deliver quantitative predictions. Unlike traditional empirical approaches, such as those described in [1], the present approach addresses complexities associated with low NO_x combustors, for example, the importance of several NO_x formation routes and the apparent variable dependence of NO_x emission on combustor pressure drop. In the case of the latter, data presented in [2] indicate a pressure drop exponent that varies between >3 and 0 (see also [3,4]).

The model described here focuses on natural gas operation. The approach, however, can be applied to other gaseous fuels and can be extended to dry low emissions (DLE) liquid fueled operation. Within this paper, the NO_x formation process is addressed, where key parameters such as the NO_x formation kinetics and the impact of unmixedness are described. Subsequently, the NO_x model is derived and evaluated using high-pressure rig data. These data are acquired at a constant pressure, preheat temperature, and pilot fuel split (14 Bara, 400°C, and 0%, respectively) and at a range of flame temperatures and flow rates.

2 NO_x Formation

The evolution of a chemically reactive species within a combustor is governed by transport as a result of convection and diffusion (including the effect of turbulence) as well as the chemical reactions that control its formation and consumption. Within this section, the NO_x chemistry is described, with focus on lean-burn low-NO_x operation. Reference [2] showed that NO_x emissions observed in high-pressure rig tests could be correlated by assuming NO_x formation under ideally premixed conditions and applying a weighting to take account of fuel-to-air ratio variations that result when the premixing is nonideal. Following these observations, this section, and subsequently, the proposed model, addresses NO_x formation within ideally premixed circumstances and the evaluation of fuel/air unmixedness.

2.1 Chemistry of NO_x Formation. NO_x formation in lean

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received June 10, 2006; final manuscript received December 22, 2006. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2006: Land, Sea and Air (GT2006), May 8–11, 2006, Bercelona, Spain. Paper No. GT2006-90333.

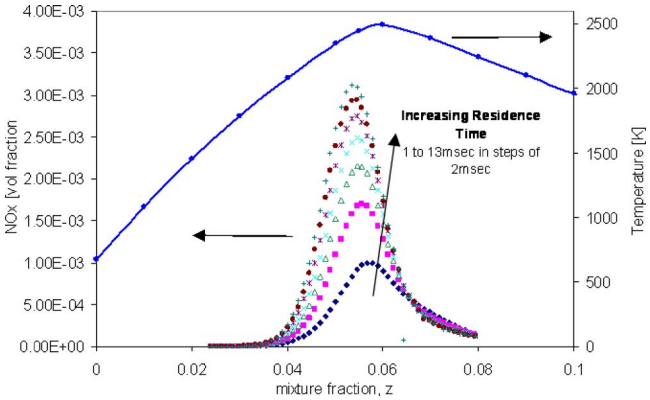


Fig. 1 PSR NO_x and adiabatic flame temperature against mixture fraction in the lean and stoichiometric regions: 14 Bara and 400°C inlet temperature. The fuel is natural gas.

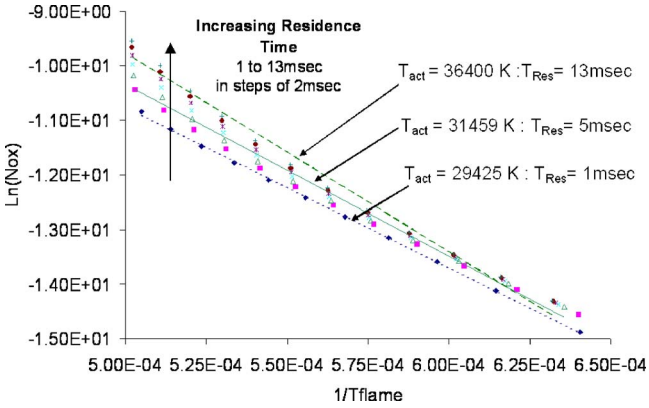


Fig. 3 PSR results: ln (NO_x) against $1/T_{flame}$ at different t_{res} including the theoretical activation temperatures: 14 Bara and 400°C inlet temperature. The fuel is natural gas.

premixed natural gas combustion at elevated pressure is thought to be dominated by the thermal and N_2O routes [5]. This notion is consistent with NO_x emissions observed in high-pressure tests [2]. Analysis of the emissions using the GRI 3.0 chemical kinetic scheme showed that other reactions involving N-H-O species, e.g., NNH, are also significant and prompt NO_x (for which HCN is the key species) had a minor influence [2].

The major parameter that affects NO_x emission is temperature. Given that NO_x formation occurs within and downstream of the reaction zone, the temperature can be related to the idealized combusted gas temperature, i.e., the adiabatic equilibrium flame temperature. Though this temperature is not exactly the temperature at which all NO_x formation occurs, it is a temperature that is readily calculated from combustor operating conditions and is thus a useful temperature to use for NO_x correlation.

Figure 1 shows the adiabatic equilibrium temperature curve around the stoichiometric region, for conditions representative of typical gas turbine operation. Also shown are perfectly stirred reactor (PSR) results of the NO_x emission at a range of PSR residence times. The GRI 3.0 reaction mechanism is used. The NO_x emissions can be seen to be confined to the higher temperature regions. The peak NO_x levels can be seen to occur slightly to the lean of stoichiometric, since oxygen is also required. The NO_x levels increase with increasing residence time. It can also be seen that the increase with residence time becomes more significant for the higher flame temperatures, in the lean region.

The impact of flame temperature on the residence time dependence of NO_x is also seen in Fig. 2. This shows the PSR NO_x

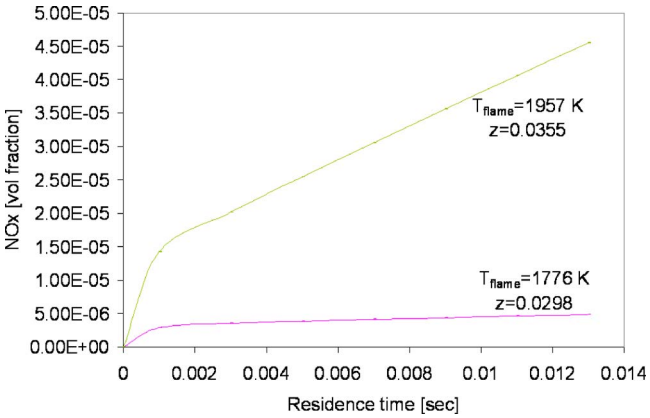


Fig. 2 PSR NO_x against t_{Res} at different mixture fractions in the lean region: 14 Bara and 400°C inlet temperature. The fuel is natural gas.

plotted against residence time for two flame temperatures in the lean region (1957 K and 1776 K). At T_{flame}=1776 K, the NO_x is seen to rapidly increase (within 1 ms) and then to vary little with residence time. In the case of T_{flame}=1957 K, there is again a rapid increase within 1 ms, and thereafter there is a continual but slower rise. This is due to two different time scales controlling NO_x emission. The slow path is due to the Zeldovich mechanism, which is highly temperature dependent and, hence, negligible at the lower T_{flame}. The fast route is due to other significant pathways, e.g., the N_2O and NNH pathways. It should be noted that there is also a fast component to Zeldovich NO_x due to the highly radical-rich environment at short residence times. The overall activation energy of the "fast" Zeldovich NO_x will differ from that of the "slow" Zeldovich NO_x, since $2O\langle=\rangle O_2$ will be far from equilibrium in the former case.

The above observations show that a suitable model for NO_x could be based on a two step process, the first step reaching equilibrium quickly, while the second step continues to rise as per the Zeldovich mechanism. The necessity to address the second step will depend on the flame temperature and the residence time within the high-temperature zone.

Figure 3 shows NO_x results from Fig. 1 on an Arrhenius-type plot, where the gradient indicates the overall activation temperature. At the shorter residence times, the data can be well approximated by a straight line, and therefore, the data can be modeled by a single Arrhenius reaction rate.

As residence time increases the magnitude of the gradient increases, due to the increasing importance of post-flame Zeldovich NO_x. Additionally, the data start to deviate from a straight line, having a greater magnitude at higher temperatures. This is again caused by the increasing dominance of the Zeldovich route. If such a plot was produced from appropriate experimental data, the characteristics of the gradient will be an indicator for the necessity of treating the Zeldovich NO_x explicitly through a two-stage NO_x model.

2.2 Effect of Unmixedness on NO_x Formation. Variation in the fuel-to-air ratio in time and space within the combustor must be considered when determining the NO_x emissions, because $NO_x(z)$ is strongly nonlinear (see, e.g., Fig. 1). Since the time and space variation of z cannot be included in a tractable model, it is most convenient to describe the variation statistically, by way of a probability density function (PDF) $P(z)$.

$P(z)$ has the property that it is bimodal with strong contributions at z=0 (100% air) and z=1 (100% fuel) in the regions where fuel is introduced. Proceeding downstream, as mixing residence time increases, $P(z)$ will resemble a Gaussian curve with its mean at the average z value for the primary zone of the combustor. As

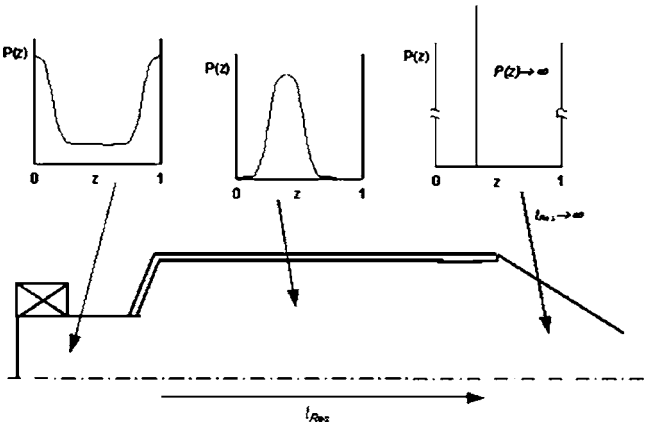


Fig. 4 Different states of mixing within the combustor

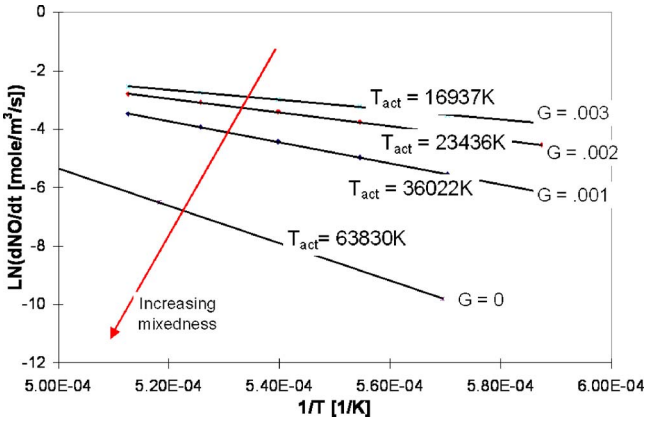


Fig. 5 Impact of unmixedness on NO_x-Zeldovich mechanism [2]

residence time increases further, $P(z)$ will become narrower, unless further inhomogeneity is introduced through, e.g., dilution jets. In the limit of infinite residence time, $P(z)$ will become a Dirac δ function at the mean value of z. This represents ideal premixing. Figure 4 illustrates the above forms. $P(z)$ can be defined by tracking its mean and variance, $\bar{z}$ and $\overline{z'^2}$. In the case of ideal premixing, $\overline{z'^2}=0$.

Typically, DLE systems will have at least two fuel streams, a main and a pilot. The former achieves a high degree of premixing for low NO_x operation, whereas the latter introduces fuel/air inhomogeneity for low load stability. The variance will be a function of the main fuel distribution and pilot fuel split. If the main fuel distribution is only a function of burner design, i.e., it is not a function of the fuel and/or airflow rates, the main fuel distribution is fixed for any particular combustor design.

Given that NO_x emission is residence time dependent, the net NO_x emission can be deduced from

$$\overline{[NO_x]} = \int\int [NO_{x_{ideal}}](z,t)P(z,t)dzdt \tag{1}$$

Section 2.1 showed the NO_x formation process can be divided into two steps, one which is relatively fast and the other slow. If it is assumed that the fast step is combustor residence time independent, then the net NO_x emission can be obtained from

$$\overline{[NO_x]} = \int [NO_{x_{fast}}](z)P(z)dz + \int_{t_1}^{t_2}\left\{\int \frac{d[NO_{x_{zel}}]}{dt}(z)P(z)dz\right\} \times(t)dt \tag{2}$$

The second step is important if the residence time and the temperature in the post-flame region are high enough for post-flame Zeldovich NO_x to be significant.

The impact of mixing quality on NO_x emission was discussed in [2]. There it was shown that as the mixing quality improves, the NO_x emissions decrease and the apparent activation temperature increases. This is in contrast to the residence time effect, where as the NO_x levels decrease (due to shorter residence time) the apparent activation temperature decreases. Figure 5 shows the impact of mixedness on the Zeldovich NO_x formation rate [2].

Figure 6 shows the results of high-pressure rig tests on two variants of a DLE combustor. These results formed the focus of [2] and are the data to be used to evaluate the presently deduced model. As discussed in [2] under conditions of near-perfect premixing, the NO_x emissions show no combustor pressure-drop dependence. In the case of nonideal premixing, there is a pressure-drop dependence, the exponent of which is not constant [2]. An effective NO_x model for DLE combustion must be able to address the variable pressure-drop dependence.

The well premixed results illustrated in Fig. 6, have been used in the present work to specify the ideal $NO_x(z)$ relationship. As discussed in [2], however, the ideal results are in agreement with PSR calculations using the GRI 3.0 mechanism. The latter could therefore also be used, in the absence of any representative well premixed data. Later in this paper, the nonideally premixed data (Fig. 6) are used to evaluate the proposed modeling concept.

Given the idealized $NO_x(z)$ relationship and noting that Zeldovich NO_x is controlled by a relatively simple well-understood mechanism, the NO_x emission can be calculated if $P(z)$ is determined in the flame and post-flame zones. As illustrated in Fig. 4, $P(z)$ varies throughout the burner/combustor. A key point, therefore, is to determine the location of the NO_x formation region, which, in turn, depends on the location of the main reaction zone.

Modification of the reaction zone location can result through variation in both the thermochemical and fluid dynamic parameters. For example, if the chemical kinetic rates increase, through, e.g., increase in preheat and/or equilibrium temperature, the mean flame location will move upstream, into a more unmixed environment. If fluid dynamic rates increase beyond critical values, through, e.g., increasing velocities, the mean flame zone will move downstream, because of an increased prevalence of local

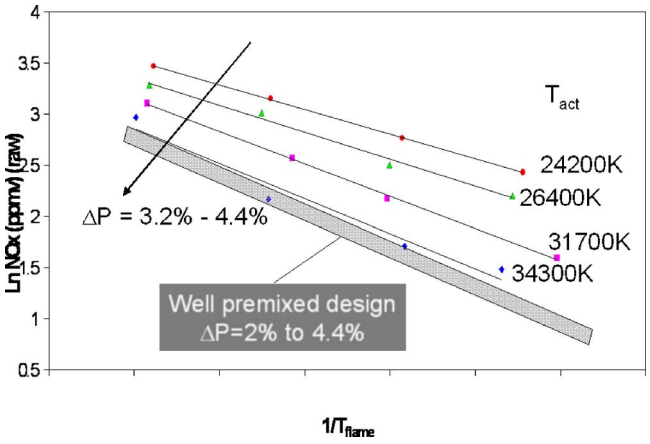


Fig. 6 NO_x emissions from high-pressure test results [2]. The data points are for the nonideally premixed configuration, and the shaded region indicates the results for the well-premixed case: P=14 Bara, inlet temperature=400°C. The fuel is natural gas.

extinction.

A convenient parameter that captures the above phenomena is the Damkoehler number (Da) (see, e.g., [6]), which is the ratio between mixing and chemical time scales

$$\mathrm{Da} = \frac{\tau_m}{\tau_c} \tag{3}$$

It should, however, be realized that both the fluid dynamics and the chemistry exhibit a large range in time scales, which may react differently to changes in global parameters. For example, fluid dynamic time scales could be defined based on the integral and Kolmogorov extremes of the turbulence spectrum or, indeed, any scale in-between. This issue is discussed further in Sec. 3.

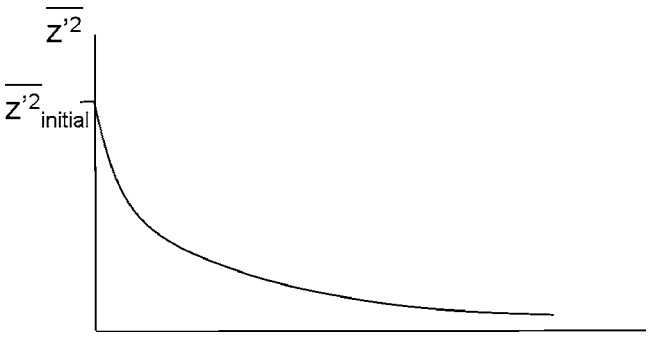


Fig. 7 Decline of the unmixedness through the combustor

3 Modelling NO_x Emission

For a typical DLE engine, the parameters that influence NO_x emission are as follows:

1. Fuel split, which impacts the mixedness
2. Volumetric flow rate, which influences residence time and turbulence levels
3. Pressure, combustor inlet temperature, and the temperature in the NO_x formation zone, which affects the chemical kinetics and the turbulence

A suitable NO_x model should therefore have these as inputs and model NO_x with sufficient accuracy. This is attempted in this section by dealing with each of the processes described in Sec. 2, namely,

1. NO_x chemistry
2. Fuel distribution
3. Heat release zone location

By way of clearly stated approximations and assumptions, we arrive at a correlation for NO_x emission. Constants in the correlation are then to be determined from experimental data. Here, we utilize the data of [2] to evaluate the feasibility of the proposed approach.

3.1 NO_x Chemistry Model. Section 2 showed that the NO_x chemistry could be split into two parts, a fast chemistry part and a slow chemistry part. The time scale of the fast chemistry processes suggests that these could be assumed to be in equilibrium and therefore are not influenced by combustor residence time. The slow chemistry process, which is increasingly dominant at high flame temperatures, exhibits a residence time dependence.

Given the analysis reported in [2], it seems that the NO_x chemistry can be fairly well represented through chemical reactor modeling using the GRI 3.0 mechanism. Correct usage of the mechanism can be enhanced by appropriate rig measurements [2].

3.2 Fuel Distribution Model. As shown earlier, the state of the fuel/air unmixedness can be described in terms of $P(z)$, which, for a given $\bar{z}$, can be described by the variance of mixture fraction $\overline{z'^2}$, which is generated through local turbulence and the injection of fuel and/or air. It subsequently decays though the action of turbulent mixing and turbulence decay. The progress of $\overline{z'^2}$ within the burner/combustor is illustrated in Fig. 7. $\overline{z'^2}$ has an initial value that is related to, e.g., fuel flow split, and subsequently decays. Eventually, $\overline{z'^2}$ will become zero, representing ideally premixed conditions. The magnitude of the initial value and the decay rate will be a function of the burner design and operating conditions.

To describe the unmixedness, therefore, a model is needed that establishes the initial value of $\overline{z'^2}$ and its subsequent decay as a function of downstream distance. It should be noted that this downstream distance is a notional distance rather than an actual distance, since it takes into account flow characteristics, such as reverse flow zones.

In order to determine $\overline{z'^2}$, we consider its conservation equation (see, e.g., [6])

$$\frac{\partial}{\partial t}\rho\overline{z'^2} + \nabla\cdot(\rho\bar{u}\overline{z'^2}) = \nabla\cdot(\overline{\rho D\,\nabla z'^2}) - \nabla\cdot(\rho\cdot\overline{u'z'^2}) - 2\rho\overline{u'z'}\cdot\nabla\bar{z} - 2\rho\overline{D\,\nabla z'\,\nabla z'} \tag{4}$$

Equation (4) assumes constant density flow, since it is assumed that the fuel/air mixing in the NO_x formation zone is dominated by the mixing that occurs within the precombustion zone (i.e., the premixing zone). Density variations due to the different densities of the fuel and air are therefore neglected. This is a poor assumption close to the fuel injector but is a reasonable assumption further downstream, since the mean density is then dominated by the air density, in the case of high heating value fuels.

A further simplification to Eq. (4) can be applied, since the flows of interest are at high Reynolds number, in which case the molecular diffusion term can be neglected because it is much smaller than the turbulent diffusion term. If closure assumptions are then applied as per the eddy-viscosity-based k-ε turbulence model [7], Eq. (4) reduces to

$$\frac{\partial\bar{\rho}\overline{z'^2}}{\partial t} + \nabla\cdot(\bar{\rho}\bar{u}\overline{z'^2}) = \nabla\cdot\left(\frac{\mu_t}{\sigma_t}\nabla\overline{z'^2}\right) + 2\frac{\mu_t}{\sigma_t}(\nabla\bar{z})^2 - c_{g_2}\bar{\rho}\frac{\varepsilon}{k}\overline{z'^2} \tag{5}$$

where the terms on the right represent turbulent diffusion, turbulence generation, and viscous dissipation, respectively.

As stated earlier, the initial value of $\overline{z'^2}$ at the head of the combustor will depend on the fuel injection characteristics. Thereafter, $\overline{z'^2}$ will decay because the dissipation term is always active, provided that $\overline{z'^2} > 0$, and the generation term will decline as gradients in the mean decline due to turbulent dispersion. In order to establish a correlation for the variation of $\overline{z'^2}$ within the combustor, therefore, it seems sensible to establish a correlation for the initial value and then for the subsequent decay.

If the transient and transport terms are neglected in Eq. (5), then the generation and dissipation terms are equal

$$c_{g_2}\bar{\rho}\frac{\varepsilon}{k}\overline{z'^2} = 2\frac{\mu_t}{\sigma_t}(\nabla\bar{z})^2 \tag{6}$$

Rearranging and replacing μ_t by its definition in terms of k and ε

$$\mu_t = c_\mu\frac{k^2}{\varepsilon} \tag{7}$$

We have

$$\overline{z'^2} = \frac{2c_\mu}{c_{g_2}\sigma_t}\frac{k^3}{\varepsilon^2}(\nabla\bar{z})^2 \tag{8}$$

Noting that $\varepsilon = k^{3/2}/L_i$, we have

$$\overline{z'^2} = \frac{2c_\mu}{c_{g_2}\sigma_t} L_i^2 (\nabla \bar{z})^2 \tag{9}$$

$\overline{z'^2}$ is therefore dependent on the integral length scale and the spatial gradient of $\bar{z}$. Now the integral length scale is proportional to the dimensions of the domain, say Φ_p the burner diameter. The expression can therefore be simplified to yield

$$\overline{z'^2_{\text{initial}}} = A_{z'} \Phi_p^2 (\nabla \bar{z})^2 \tag{10}$$

where $A_{z'}$ is a constant that must be determined through correlation of experimental data.

$(\nabla \bar{z})$ is related to the fuel distribution. If premixing occurs instantaneously, $(\nabla \bar{z})$ is zero and, therefore, so is the initial $\overline{z'^2}$. $(\nabla \bar{z})$ should be dependent on the fuel split. In the case of zero pilot fuel flow, $(\nabla \bar{z})$ may be zero, if the main fuel results in ideal premixing being achieved well ahead of the flame zone. This is the case for the well premixed case reported in [2].

$(\nabla \bar{z})$ needs to be determined for each combustor at a reference point, e.g., zero pilot operation. A suitable model is then required to reflect its dependence upon fuel split. It could be expressed as

$$(\nabla \bar{z}) \propto \frac{(\Delta \bar{z})}{\Delta x} \tag{11}$$

If Δx is assumed to be proportional to the dimensions of the domain, then we could relate it to, e.g., the burner diameter, leading to

$$(\nabla \bar{z}) = A_z \frac{z_{\text{diff}}}{\Phi_p} \tag{12}$$

At the simplest level, the front end of the combustor could be divided into two sections, an inner part and an outer part. Each part could be assigned a value of $\bar{z}$, and z_{diff} is then simply the difference of the assigned values, leading to

$$\nabla \bar{z} = A_z \frac{z_{\text{outer}} - z_{\text{inner}}}{\Phi_p} \tag{13}$$

The impact of the fuel streams on the value of z in the inner and outer regions, should then reflect any particular configuration. For example, if in the case of a two-fuel stream system (say a main and a pilot stream) the pilot is configured to support flame stabilization along the burner axis, then the pilot fuel split should increase z_{inner} and decrease z_{outer}.

Within the model, it is important only to establish the impact of any change in burner operation on z_{diff}. The absolute value of z_{diff} is not important given that A_z is to be determined empirically.

The decay of $\overline{z'^2}$ with downstream distance can be estimated from Eq. (5), where the transient term is neglected (steady state) and turbulent diffusion is neglected, as this serves only to redistribute $\overline{z'^2}$. If it is further assumed that during the decay process the generation term is negligible compared to the dissipation term, Eq. (5) reduces to

$$u \frac{d(\overline{z'^2})}{dx} = -c_{g_2} \frac{\varepsilon}{k} \overline{z'^2} \tag{14}$$

In Eq. (14), the system is reduced to one spatial dimension, x.

Rearranging Eq. (14) and integrating leads to

$$\int_{x_{\text{init}}}^{x} \left(\frac{\overline{dz'^2}}{\overline{z'^2}} \right) = -c_{g_2} \int \frac{\varepsilon}{k} \frac{1}{u} dx \tag{15}$$

Assuming ε and k to be integral properties that do not vary with x, assuming u to be a bulk velocity in the burner and noting that ε/k is the turbulence integral strain rate, we have

$$Ln\left(\frac{\overline{z'^2}}{\overline{z'^2_{\text{init}}}} \right) = -c_{g_2} \frac{u'}{L_i} \frac{1}{\bar{u}} x \tag{16}$$

For high Reynolds number, u' is directly proportional to a bulk average velocity and the turbulent integral length scale is proportional to a combustor dimension, say Φ_p, which gives

$$Ln\left(\frac{\overline{z'^2}}{\overline{z'^2_{\text{init}}}} \right) = -c_{g_2} \frac{T_i \bar{u}}{A_{L_i} \Phi_p} \frac{1}{\bar{u}} x = -\frac{c_{g_2} T_i}{A_{L_i}} \frac{1}{\Phi_p} x \tag{17}$$

where T_i is the turbulence intensity $u'/\bar{u}$.

Lumping all the constants together and taking the exponential of both sides, yields

$$\left(\frac{\overline{z'^2}}{\overline{z'^2_{\text{init}}}} \right) = e^{-A_{z/2}(1/\Phi_p)x} \tag{18}$$

The fuel unmixedness is therefore defined in terms of an initial value, which, for a given combustor, is dependent on burner operation, e.g., fuel flow split, and the exponential decay of the unmixedness with distance.

The next step is to establish a model for the heat release process, in particular, the impact that thermochemical and fluid dynamic parameters have on its mean location and extent in x space. This would then allow the level of unmixedness in the main NO_x forming zone to be determined.

3.3 Heat Release Location Model. The location of the heat release zone will be modified based on the relative magnitudes of the chemical kinetic rate and the fluid dynamic strain rate, since this affects the degree of local flame extinction. In other words, if strain rate increases, leading to increased local extinction, the heat release zone requires a greater volume for the fuel to be consumed, and its mean location will move downstream.

The Damkoehler number, introduced in Eq. (3), seems a logical parameter to indicate the relative importance of chemical kinetic and strain rates and, hence, the reaction zone location in x space.

Given that turbulence exhibits scales from the integral to the Kolmogorov scales, a number of strain rates can be determined. Next, we confine attention to the two extremes. Additionally, we assume that the turbulence within the flame zone is dominated by the turbulence in the approach flow just ahead of the flame.

The integral strain rate can be defined as

$$a_i = \frac{u'}{L_i} \tag{19}$$

u' can be related to a bulk velocity and the integral length scale to the burner diameter

$$a_i = \frac{T_i \bar{u}}{A_{L_i} \Phi_p} = \frac{T_i}{A_{L_i}} \frac{1}{\Phi_p} \frac{M}{\rho_u \frac{\pi}{4} \Phi_p^2} = \frac{4 T_i}{A_{L_i} \pi} \frac{M}{\rho_u \Phi_p^3} = A_{a_i} \frac{M}{\rho_u \Phi_p^3} \tag{20}$$

Equation (20) shows that, for a particular combustor, the integral strain rate is proportional to the burner mass flow M divided by the unburnt density, which is assumed to be that of the combustor inlet flow, since the velocity is based on the cold flow inlet velocity.

The Kolmogorov strain rate a_η can be expressed in terms of integral scale parameters

$$a_\eta = \frac{u'^{3/2}}{\nu^{1/2} L_i^{1/2}} = \frac{T_i^{3/2}}{A_{L_i}^{1/2}} \frac{\bar{u}^{3/2}}{\nu^{1/2} \Phi_p^{1/2}} \tag{21}$$

With the above equations, the Kolmogorov strain rate can be expressed as follows:

$$a_\eta = \frac{T_i^{3/2}}{A_{L_i}^{1/2}} \frac{1}{\nu^{1/2} \Phi_p^{1/2}} \left(\frac{M}{\rho_u \frac{\pi}{4} \Phi_p^2} \right)^{3/2} = A_\eta \frac{M^{3/2}}{\nu^{1/2} \rho_u^{3/2} \Phi_p^{7/2}} \tag{22}$$

The Kolmogorov strain rate is proportional to mass flow to the

power 3/2.

An approximation for the dynamic viscosity of dry air is

$$\mu = \mu_K T^{0.6} \tag{23}$$

μ_K is a dimensional constant of value 0.65×10^{-6} kgm^{-1} s^{-1} K$^{-0.6}$. Equation (23) is a good fit for the data tabulated in [8]. Substituting this into Eq. (22), yields

$$a_\eta = A_\eta \frac{M^{3/2}\rho_u^{1/2}}{\mu_K^{1/2}T^{0.3}\rho_u^{3/2}\Phi_p^{7/2}} = A_\eta \frac{M^{3/2}}{\mu_K^{1/2}T^{0.3}\rho_u\Phi_p^{7/2}} \tag{24}$$

Integral and Kolmogorov time scales can be determined, since they are simply the reciprocal of a_i and a_η, respectively,

$$\tau_{mi} \propto \Phi_p^3 \left(\frac{M}{\rho_u}\right)^{-1} \tag{25}$$

$$\tau_{m\eta} \propto \nu^{1/2}\Phi_p^{7/2}\left(\frac{M}{\rho_u}\right)^{-(3/2)} \tag{26}$$

Equations (25) and (26) show that different scales within the turbulence spectrum have different dependencies on the global combustor operating parameters. Therefore, the way in which the mean heat release zone location changes with combustor operating parameters depends on the scales of turbulence with which it most interacts.

Let y represent the turbulence scale for which the timescale is equal to the chemical time scale τ_c. If it is assumed that y is at an intermediate scale between the integral and Kolmogorov scales, then making the usual assumption of turbulence dissipation being rate limited by the largest scales and noting that the length and velocity scale are related to the appropriate time scale as follows:

$$\frac{L_y}{u_y} = \tau_y \tag{27}$$

setting τ_y equal to τ_c gives

$$L_y = \frac{\tau_c^{3/2}u'^{3/2}}{L_i^{1/2}} \tag{28}$$

Because y, by definition, is the scale that interacts with the heat release reactions, we assume that change in L_y is directly related to the distance parameter x identified in Eq. (18). Such a notion seems appropriate for recirculation-zone stabilized combustion, where the leading edge of the reaction zone is fixed by the recirculation zone. The mean reaction zone location can then only change due to variation in the mean reaction zone thickness. The assumption here is that this thickness is proportional to L_y.

If we reintroduce the Damkoehler number (Eq. (3)), where the mixing time scale is given by the integral timescale (the reciprocal of Eq. (19)) Eq. (28) becomes

$$L_y = \mathrm{Da}^{-(3/2)}L_i \tag{29}$$

Noting the assumed proportionality between L_y and x and that L_i is proportional to the combustor dimension, Eq. (29) can be substituted into Eq. (18) to yield

$$\left(\frac{\overline{z'^2}}{\overline{z'^2_{\mathrm{init}}}}\right) = e^{-A\mathrm{Da}^{-(3/2)}} \tag{30}$$

where A is a constant that needs to be determined by correlation of experimental data. The Damkoehler number also needs to be determined, in particular, the specification of the chemical time scale.

The model is now complete and can yield the NO_x emissions in the following steps:

1. Determine Da from the combustor operating conditions
2. Calculate the variance of mixture fraction from Eq. (30)
3. Calculate the NO_x emission from Eq. (2)

Section 3.4 investigates the suitability of the approach through conducting a preliminary evaluation against test rig data.

3.4 Evaluation of the Model. Using the definition of Da (Eq. (3)) and using Eq. (25) for the mixing time scale, Eq. (30) can be written in terms of the mass flow rate and combustor inlet density

$$\left(\frac{\overline{z'^2}}{\overline{z'^2_{\mathrm{init}}}}\right) = e^{-A(1/\Phi_p^{9/2})\tau_c^{3/2}(M/\rho_u)^{3/2}} \tag{31}$$

Let us suppose that the chemical time scale is given by the characteristics of an appropriate laminar premixed flame. Such a notion implies that combustion occurs within the laminar flamelet regime (see, e.g., [6]). This is questionable within lean premixed gas turbine combustion, which most likely occurs within the thickened flamelet regime. Such issues should be investigated further should the end expression prove inappropriate for correlating against experimental data.

The chemical time scale for a laminar premixed flame can be expressed as (see, e.g., [6])

$$\tau_c \propto \frac{\nu}{S_l^2} \tag{32}$$

Equation (32) in addition to the approximations described in [6], assumes that the Prandtl/Schmidt number is constant. S_l can be determined through measurement or computation of freely propagating laminar flames.

Next, we look at the ability of the deduced expression to correlate test rig data that had been gathered at a single pressure and preheat temperature and at a range of flame temperatures. In such a case, it is not unreasonable to assume that S_l is a function of temperature (T_{flame}) only. Since the laminar flame speed will reach a maximum at or near stoichiometric conditions, it is not unreasonable to assume that S_l is proportional to temperature to a power less of <1 as stoichiometric conditions are approached, and ~ 1 over a limited extent of the lean region. If a linear relationship is substituted into Eq. (32), and Eq. (23) is used, we have

$$\tau_c \propto \frac{T_{\mathrm{flame}}^{0.6}}{\rho_{\mathrm{flame}}T_{\mathrm{flame}}^2} \propto \frac{T_{\mathrm{flame}}^{0.6}T_{\mathrm{flame}}}{T_{\mathrm{flame}}^2} = \frac{1}{T_{\mathrm{flame}}^{0.4}} \tag{33}$$

Given the assumptions used to establish the exponent of T_{flame}, the value 0.4 must be considered as approximate, and though a realistic range can be determined a priori, its value should be established from experimental observations. Substituting Eq. (33) into Eq. (31), results in $\tau_c^{3/2}$ of Eq. (31) being replaced by a proportionality to $T_{\mathrm{flame}}^{-0.6}$. If, however, rather than assuming a linear dependence of S_l on T_{flame} (as in Eq. (33)) one assumes S_l to be proportional to $T_{\mathrm{flame}}^{0.9667}$, Eq. (31) would become

$$\left(\frac{\overline{z'^2}}{\overline{z'^2_{\mathrm{init}}}}\right) = e^{-A'T_{\mathrm{flame}}^{(1/2)}(M/\rho_u)^{3/2}} \tag{34}$$

The T_{flame} exponent is therefore very sensitive to the S_l characteristics. However, evaluation of the impact of the T_{flame} exponent on the end results of the model, shows it to have a relatively minor influence, when utilizing the present experimental data set [2]. This issue will be the focus of further work, where data will be acquired over a wider range of conditions. In the work reported below, we have adopted the $-1/2$ exponent as shown in Eq. (34). Additionally, we will here confine our attention to a single combustor. The burner dimension parameter has therefore been incorporated into the empirical constant A' in Eq. (34).

The data reported in [2], which will be used for the evaluation of the modeling concept, did not illustrate a residence time dependence [2]. Post-flame NO_x generation can therefore be neglected, and Eq. (2) can be simplified to

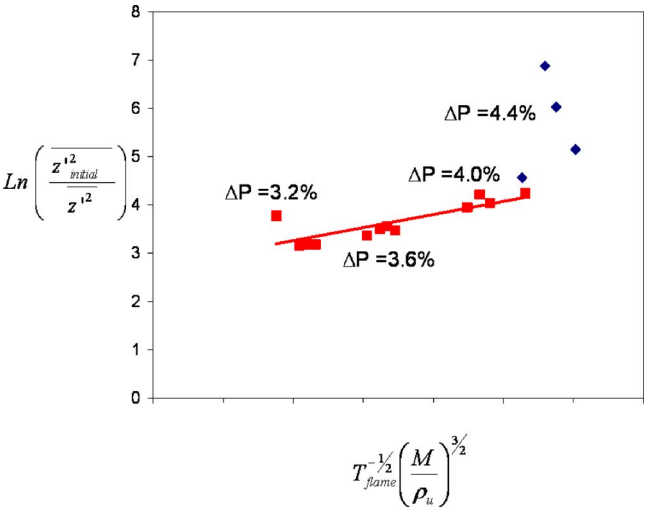


Fig. 8 Analysis of Eq. (34)

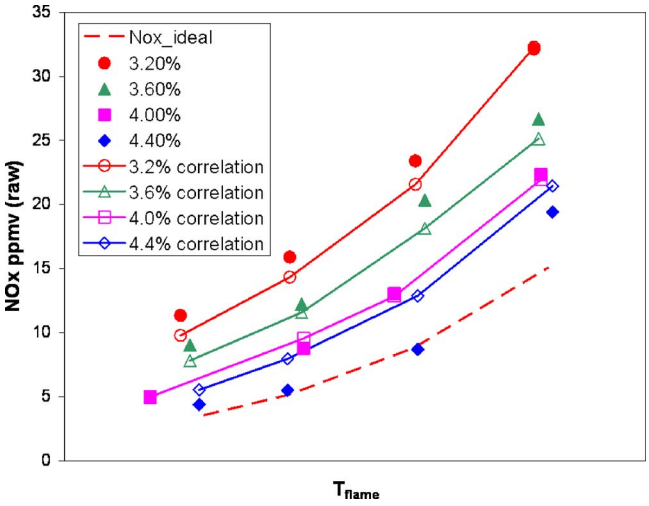


Fig. 9 Comparison between NO_x from rig tests [2] and model (Eqs. (34) and (35))

$$\overline{[NO_x]} = \int [NO_{x_{fast}}](z)P(z)dz \tag{35}$$

Using the well premixed results in [2] to establish $[NO_{x\ fast}](z)$, and utilizing the nonideally premixed results in [2] as the base NO_x data (i.e., the left-hand side of Eq. (35)), the only unknown is $P(z)$. If this is presumed to be a β function, as is common practice (see, e.g., [6]), then noting that the mean value of z is given by the primary zone fuel-to-air ratio, the only unknown is $\overline{z'^2}$. The latter can thus be deduced through inversion of Eq. (35).

If the initial value of $\overline{z'^2}$ is specified, the left-hand side of Eq. (34) can be established. In the data from [2], no variation in fuel flow spilt is considered, and therefore as discussed there, there is negligible variation in fuel distribution as the burner operating parameters (mass flow and fuel-to-air ratio) are varied.

Given the $\overline{z'^2}$ indicated by the NO_x data and having established $\overline{z'^2_{initial}}$ by fixing the main fuel distribution parameters, the validity of Eq. (34) can be evaluated. For the equation to be valid, a plot of

$$Ln\left(\frac{\overline{z'^2_{init}}}{\overline{z'^2}}\right) \quad \text{against } T_{flame}^{-(1/2)}\left(\frac{M}{\rho}\right)^{3/2}$$

should yield a straight line. The gradient of that line will be A'. Figure 8 shows such a plot.

Figure 8 shows that the data, apart from some points at the highest pressure drop, are well represented by Eq. (34). The deviation at the highest pressure drop (4.4%) could be due to a substantial change in the flame characteristics, e.g., a shear layer that had sustained combustion at the lower pressure drops may have extinguished. Further investigation is required to quantify this phenomenon. The present model is unable to address such gross changes in burning characteristics. Should such a phenomenon be important, an extension to the model will be required, through the introduction of an additional parameter based on say, a Karlovitz number (see, e.g., [6]).

Figure 9 shows the end result from the model, where the NO_x emission is plotted against flame temperature for a range of pressure drops. Also shown are the experimental data.

Figure 9 shows that the model, utilizing the deduced exponent on mass flow of 3/2, represents the data very well, within 2 ppm. This is within the measurement uncertainty. The exceptions are the two central points at the highest pressure drop. These points can be seen to lie on the idealized NO_x curve, indicating that the unmixedness has dropped to a negligible value. As stated earlier, this may be due to a fundamental change in the flame, which cannot be addressed by the simple model derived here. An additional parameter related to extinction effects, however, may be sufficient to address this issue and should be a focus of further work.

4 Conclusions

The framework for a NO_x emissions algorithm has been established, which is capable of being both accurate and tractable for use within combustor design codes and on-line predictive emissions tools. The model is derived through simplification of the fundamental processes controlling NO_x emissions, whereas quantitative accuracy is obtained through incorporation of suitable empirical data. NO_x emission is modeled in two parts:

1. NO_x characteristics under ideally premixed circumstances
2. Statistics of the fuel-to-air ratio within the NO_x forming zone

Evaluation utilizing high-pressure rig data (obtained at a constant pressure and preheat temperature (14 Bara and 400°C) showed the model to perform very well. In particular, the dependence of the fuel-to-air ratio statistics on the burner flow parameter are consistent with the theoretically derived dependence. Quantitatively, the NO_x model results agree with the rig data within 2 ppm, which is within the experimental uncertainty.

A key anomaly in the model has been observed in the case of two of the test points for the highest pressure drop tested. These points indicate a major change in behavior of the flame that is not picked up in the model. This anomaly might be caused by quenching of a shear layer, or part of a shear layer that had supported combustion for the bulk of the points. This phenomenon requires further evaluation through rig testing. A model extension will be required to address this. A parameter based on a Karlovitz number seems the most appropriate model extension.

In the derivation of the empirical model, the assumptions and simplifications are clearly stated. These are important to note, since some simplifications that can be made in the case used for the model evaluation may not be suitable for other cases.

The present modeling approach is superior for low-emission applications to those described in [1]. There is clearly an improvement in the description of the NO_x chemistry that is appropriate to lean premixed combustion. More significantly, however, though it is recognized in many previous models (see [1]) that mixing affects NO_x emission, this had been introduced by a simple $(\Delta P/P)^n$ or mass flow-rate term This is sufficient to account for the effect of mixing quality on NO_x when the NO_x emission is large and interest is limited to a relatively small range in operating

parameters; however, it is a poor description for lean premixed combustion. In these systems, as the ideally premixed condition is approached, the pressure-drop dependence varies significantly, reaching a zero dependence in the ideally premixed case, as we ([2]) and others (e.g., [3,4]) have found. The present modeling approach treats the fluid dynamics impact and the fuel/air unmixedness explicitly. If the present model were described in terms of $(\Delta P/P)^n$, one would see that the exponent would be variable. The exponent is also capable of reaching 0, showing no $(\Delta P/P)$ dependence, in the ideally premixed case. This, the model is able to do without any artificial cutoff or blending terms added.

Acknowledgment

Kirsten Roden's role in the present work was in support of her diploma studies. She is pleased to acknowledge the valuable support and guidance given to her by her academic supervisors, Professor J. Seume and S. Kanzer of the University of Hannover, Germany.

Nomenclature

A = constant
A' = constant
DLE = dry low emissions
FAR = fuel-to-air ratio by mass
G = $\overline{z'^2}/[\bar{z}(1-\bar{z})]$
K = turbulence kinetic energy
L = length scale
M = burner mass flow
PDF = probability density function
$P(z)$ = PDF of z
PSR = perfectly stirred reactor
T_{act} = activation temperature
T_{flame} = adiabatic equilibrium primary zone temperature
T_{res} = residence time
z = mixture fraction=FAR/(1+FAR)
ΔP = pressure drop
ε = turbulence dissipation rate
$\bar{\phi}$ = average of ϕ
ϕ' = fluctuation of ϕ
Φ_p = burner diameter
μ = dynamic viscosity
ν = kinematic viscosity
ρ = density
τ = time scale

Subscripts

c = chemical
flame = equilibrium conditions
I = integral
init = initial
m = mixing
u = unburnt
η = Kolmogorov

References

[1] Lefebvre, A. H., 1998, *Gas Turbine Combustion*, 2nd ed., Taylor and Francis, New York.
[2] Syed, K. J., and Buchanan, E., 2005, "The Nature of NO_x Formation Within an Industrial Gas Turbine Dry Low Emission Combustor," ASME GT Expo, Reno, June, ASME Paper No. GT2005–68070.
[3] Mansour, A., and Benjamin, M. A., 2003, "Emission Performance of the Parker Macrolaminate Premixer Tested Under Simulated Engine Conditions," ASME Paper No. GT2003–38010.
[4] Leonard, G., and Stegmaier, J., 1993, "Development of an Aeroderivative Gas Turbine Dry Low Emissions Combustion System," ASME Paper No. 93-GT-288.
[5] Nicol, D. G., Steel, R. C., Marinov, N. M., and Malte, P., 1993, "The Importance of the Nitrous Oxide Pathway in Lean-Premixed Combustion," ASME Paper No. 93-GT-342.
[6] Williams, F. A., 1985, *Combustion Theory: The Fundamental Theory of Chemically Reacting Flow Systems*, 2nd ed., Benjamin Cummins, Menlo Park, CA.
[7] Jones, W. P., and Launder, B. E., 1972, "Prediction of Relaminarization With a Two-Equation of Turbulence Model," Int. J. Heat Mass Transfer, **15**, pp. 301–314.
[8] Mayhew, Y. R., and Rogers, G. F. C., 1967, *Thermodynamic and Transport Properties of Fluids—SI Units*, 2nd ed., Blackwell and Mott, Oxford, UK.

Klaus Peter Geigle
e-mail: klauspeter.geigle@dlr.de

Wolfgang Meier

Manfred Aigner

Institute of Combustion Technology,
German Aerospace Center (DLR),
Pfaffenwaldring 38-40,
D-70569 Stuttgart, Germany

Chris Willert

Marc Jarius

Institute of Propulsion Technology,
German Aerospace Center (DLR),
Linder Höhe,
D-51147 Köln, Germany

Patrick Schmitt
CERFACS,
42 Avenue Gaspard Coriolis,
F-31057 Toulouse Cedex 1, France

Bruno Schuermans
ALSTOM Power Ltd.,
Brown Boveri Str. 7,
CH-5401 Baden, Switzerland

Phase-Resolved Laser Diagnostic Measurements of a Downscaled, Fuel-Staged Gas Turbine Combustor at Elevated Pressure and Comparison to LES Predictions

A technical gas turbine combustor has been studied in detail with optical diagnostics for validation of large-eddy simulations (LES). OH^* chemiluminescence, OH laser-induced fluorescence (LIF) and particle image velocimetry (PIV) have been applied to stable and pulsating flames up to 8 bar. The combination of all results yielded good insight into the combustion process with this type of burner and forms a database that was used for the validation of complex numerical combustion simulations. LES, including radiation, convective cooling, and air cooling, were combined with a reduced chemical scheme that predicts NO_x emissions. Good agreement of the calculated flame position and shape with experimental data was found. [DOI: 10.1115/1.2718222]

Keywords: gas turbine combustor, laser diagnostics, thermoacoustics, numerical simulation

Introduction

Thermoacoustic instabilities are an important issue in the design and operation of low-emission gas turbine combustors and thus one focus of research activities [1–4]. One approach for reduction of instabilities while remaining to have low emissions of NO_x is the use of fuel-staged burners. Independent tuning of both fuel lines enables a continuous variation of the flame between a pure diffusion flame and a quasi-pre-mixed combustion at constant total fuel flow. This feature shall be used to simultaneously minimize pulsations and NO_x emissions. In the investigation presented, laser diagnostic as well as conventional measuring techniques were applied to characterize the combustion under different operating conditions. A major aspect was the visualization of the varying flame structures in pulsating flames by phase-locked measuring techniques.

Regions of heat release as well as flame structures were studied by applying OH^* chemiluminescence imaging and OH laser-induced fluorescence (LIF). Two different combustion modes have been detected: One with the flame attached to the burner and the other with a lifted flame. For selected operating conditions the pressure signal from the combustion chamber was used for triggering phase resolved measurements revealing changes of the flame structure during a cycle of the instability. Particle image velocimetry (PIV) applied to some of the flames resulted in additional information about the flow field in stable and pulsating flames. Also, the NO_x emission, the exhaust gas temperature, and the wall temperatures were determined by probe techniques and thermocouples.

Experiment

Burner and Flames. The burner studied was a not optimized downscale of a recently developed full-scale industrial burner for operation in a gas turbine [5]. It consists of two half cones, which are shifted radially apart to give way to two air inlet slots with constant slot width. A central lance induces additional air in the axial direction of the burner. Natural gas is injected into the preheated air flows through a number of holes along the inlet slots and the central lance (Fig. 1). The staging parameter α is defined as the ratio of lance fuel relative to total fuel. As the flow exits to the combustion chamber, inner and outer recirculation zones are formed to stabilize the lean partially premixed flame in varying distance from the burner.

For optical diagnostics, the pressure vessel and the combustion chamber were equipped with large quartz windows, the outer ones being exposed to the pressure and intermediate to low temperature, the inner ones to high temperature but only moderate pressure gradients.

A fraction of the 673 K hot preheated air is used for impingement cooling of the burner front panel and to provide a film cooling of the windows, additionally preventing too strong pollution by PIV seeding particles. Leakage through the burner front panel and for the window film as well as system leakage were determined separately and taken into account. Nevertheless, these cool-

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received April 21, 2006; final manuscript received September 19, 2006. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2006: Land, Sea and Air (GT2006), May 8–11, 2006, Barcelona, Spain. Paper No. GT2006-90875.

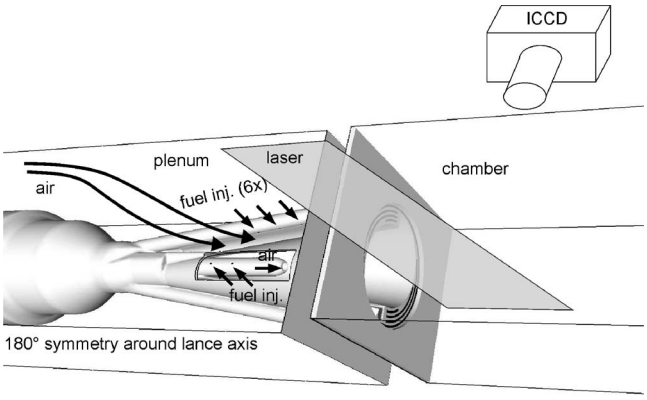


Fig. 1 Burner and optical setup

ing contributions are not representative for full-size industrial burners; therefore, reported quantities might not be transferable to the technical case one to one.

Acoustic Setup. The experimental setup for the presented phase-resolved OH measurements consists of an optical component and a triggering unit based on the measured dynamic pressure. The dynamic pressure trace was used rather for triggering purposes than for quantification of the oscillations. For the determination of the dynamic pressure, the combustion chamber was equipped with a pressurized microphone (probe microphone 4182, Brüel & Kjaer) connected to the downstream part of the combustion chamber and a pressure transducer (Type 4043A10, Kistler) close to the burner front panel. Both devices typically showed a similar behavior, so the signal from either of them could be used as an external trigger of a delay generator (DG535, Stanford Research Systems) after adequate filtering. The output of this unit, delayed in steps of 45 deg, served as the input trigger for the intensified charge-coupled device (ICCD) and for the laser exciting OH LIF [6].

OH LIF and Chemiluminescence. A line-of-sight integrated OH^* emission signal was detected using an image-intensified CCD camera (PI Max 512 UV, Princeton Instruments) equipped with an achromatic $f/2$ UV lens with 100 mm focal length (B. Halle Nachfl.) and an interference filter centered at 312 nm with a 20 nm bandwidth. The optical setup allowed for imaging of the complete flame; the alignment of laser and detection relative to the burner is illustrated in Fig. 1. The integration time for collecting sufficiently high signal levels was set to 32 μs, thus clearly below the time scale of the periodic fluctuations. Images were corrected for background and the sensitivity distribution of the CCD camera.

For OH LIF measurements, a 50 mm wide laser light sheet was introduced into the combustion chamber perpendicular to the detection path. Therefore, the 532 nm output of a pulsed Nd:YAG laser was used to pump a dye laser. After frequency doubling, the overlapping $Q_1(9)$ and $Q_2(8)$ rotational lines of the $A\,^2\Sigma^+ - X\,^2\Pi$ ($\nu'=1$, $\nu''=0$) OH transition were excited at 284.84 nm with a laser power of $\sim$3 mJ/pulse. A beam splitter before the combustion chamber reflected part of the incident laser beam into a quartz cell containing dye solution to allow for correction of the acquired LIF images with the instantaneous laser intensity distribution.

PIV. The flow field was monitored with particle image velocimetry (PIV). This planar light-sheet technique is based on measuring the displacement of tracer particles using a pair of images that were recorded in short succession [7,8]. Because of the high-temperature environment the use of solid tracer materials was essential to prevent evaporation of the tracers. In this application, ground alumina (Al_2O_3) was chosen. An average size of <1 μm ensured good flow response. A divergent light sheet of 1 mm thickness was formed using a system of one spherical and two cylindrical lenses, and introduced to the combustion chamber in place of the OH LIF optics to measure the axial (u) and radial (v) velocity components. A standard, frequency-doubled, double-cavity, pulsed Nd:YAG laser (NewWave, Gemini PIV) with a pulse energy of 120 mJ per pulse at 532 nm was used for tracer illumination. The pulse delay between the two laser pulses was $\tau=5$ μs. On the detection side, a thermoelectrically cooled, interline transfer CCD camera (PCO, 1280×1024 pixel resolution) with a $f/2.8$ lens of 55 mm focal length (Nikon) was used to capture image pairs with very short interframing times (<1 μs) at recording rates of up to four image pairs per second. A band-pass filter with center frequency of 532 nm and 5 nm width and a shutter placed in front of the lens rejected most of the flame radiation [8]. Sequences of up to 100 image pairs were recorded to calculate the mean value of the absolute velocity. Because of the rapid rate of seeding deposition on the combustion chamber windows, only several tens of images could be acquired. Phase-resolved measurements could not be performed.

Modeling

The LES code is a fully unstructured compressible finite volume/element code, including species transport and variable heat capacities [9]. Centered spatial schemes and explicit time-advancement are used to control dissipation and capture acoustics [10]. For the present simulations, Smagorinsky's LES model [11] and the Lax-Wendroff scheme are used. Characteristic boundary conditions [12,13] are used at all boundaries.

The chemistry used for all simulations is a three-step reduced mechanism for methane/air combustion [4]. Flame/turbulence interaction for the subgrid level is modeled by the artificially thickened flame model [14]. A law of the wall approach is used for convective heat transfer at walls [15] and a Stefan-Boltzmann-type model is used to evaluate radiation effects [16].

Results and Discussion

Technical Data. The main parameters to characterize the flames are pressure p, fuel staging α, adiabatic flame temperature T(ad), including air inlet temperature and equivalence ratio, and burner exit velocity. During the experiments, the general behavior of full-scale standard burners was reproduced. Two different combustion modes were detected. In the first one, the flame stabilized inside of the burner; in the second one, it was clearly lifted and exhibited pulsations. At low adiabatic flame temperatures, the burner could not be properly stabilized and lean blow-off (LBO) occurred. One main stabilization mechanism in standard burners was probably suppressed due to requirements of large optical access and thus cold components. Conventionally, flame stabilization is improved by increasing burner front panel temperature, resulting in a prominent outer hot exhaust gas recirculation zone. This was not feasible in the current experiments.

As expected, NO_x values decrease with lower adiabatic flame temperatures at constant staging, pressure, and burner exit velocity (trends shown in Fig. 2, left, for different p and α). During the experiments, NO_x was a good indicator for imminent flame blow off; when approaching one-digit values, instabilities leading to flame extinction occurred.

Varying the staging while keeping the total fuel mass flow constant resulted in strong changes of exhaust NO_x. For high α (high fuel flow through the burner lance), the values dramatically increased, indicating a worse fuel-air mixture. At low staging, variations in NO_x formation were less pronounced, showing a similar minimum for different p and T(ad) (see Fig. 2, right). The influence of pressure and burner exit velocity on NO_x formation (shown in Fig. 3) indicates an increase of NO_x with pressure and a decrease with increasing burner exit velocity for the studied low staging values. To conclude, the downscaled burner, mounted in a

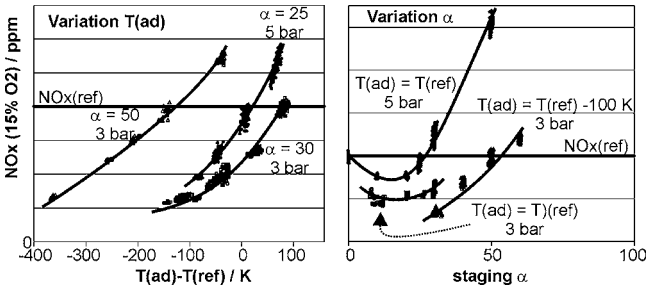


Fig. 2 NO_x value dependency on adiabatic flame temperature for different pressures and staging (left) and on staging keeping the other parameters constant (right).

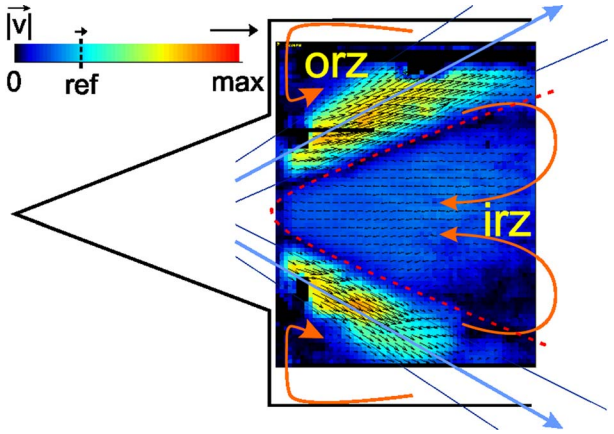


Fig. 4 Averaged flow field measured with PIV for a stable flame at 3 bar, intermediate α and corresponding recirculation zones in the flame.

combustion chamber specially designed for optical experiments at increased pressure seems to represent technical engine burners quite well.

Flow Field. The particle imaging velocimetry experiments turned out to be quite challenging. In spite of the implementation of a window film for cooling and especially preventing the windows from being coated by seeding particles, the duration of experiments for PIV was extremely short. Even if seeding was only applied during the seconds of the PIV experiments (thus after combustor stabilization), the windows rapidly became coated with particles. During initial test runs, visualization of the seeding particles was not at all successful. Finally, the influence of laser light reflections in the combustion chamber could be minimized and seeding under the chosen harsh conditions be optimized, resulting in several sequences of velocity images. Pulsating and stable flames were studied at the same burner exit velocity up to 8 bar. Because of rapid window coating coupled with the long cycle times involving cool-down sequence, window cleaning, and subsequent facility restart and stabilization, phase-resolved measurements of the pulsating flames were not conducted.

With regard to measurement precision, the noise on displacement measurement obtained from a 32×32 pixel sampling window (3×3 mm^2) is on the order of 0.05 pixel, which corresponds to ~1 m/s in the flow field. This could have been improved either by increasing the time delay between the two laser pulses or by increasing the sampling area over which the PIV displacement is computed (e.g., averaged). The former approach comes at the cost of signal loss due to possible out-of-plane motion of the particles (the flow has a strong axial swirl), whereas the increase of sampling areas would result in a proportional reduction in spatial resolution. The measurement uncertainty of the averaged flow field reduces proportional to the square root of the number of averaged samples: For a 100 image sequence, the measurement uncertainty then is on the order of 0.1 m/s.

One typical average velocity field for a stable 3 bar flame is shown in Fig. 4, demonstrating the main features of this type of flame. Relative high-speed fresh gas (red) is injected into the combustion chamber on the order of twice the calculated burner exit velocity (that is based on the full burner exit area) with an opening angle of $\sim$58 deg. This region is separated from the wide inner recirculation zone (IRZ, darkgray (dark blue), axial velocity $u\approx-0.4v(\mathrm{ref})$), transporting hot exhaust gas back to the burner, responsible for ignition of the injected fresh gas in the shear layer between the counterflowing volumes. The stagnation point is very close to the burner exit on the axis. From PIV images, an outer recirculation zone (ORZ) can only be estimated. Close to the windows and their mounting posts, laser light scattering from surfaces prevented a proper data analysis. These positions are masked in the images. The existence of the ORZ is better visualized by the OH LIF data presented in the following section.

In Fig. 5, the influence of pressure on the flow field is demonstrated. A crack of the combustor window is visible at 5 bar (middle) and 8 bar (right), probably due to frequent system shutdown during the PIV experiments and corresponding strong changes of combustor load. The averaged PIV images of pulsating flames do not show significant differences relative to the stable flame (Fig. 4), except a moderately faster backflow in the IRZ (axial velocity $u\approx-0.5v(\mathrm{ref})$). Additionally, the 8 bar flame shows a wider opening angle (70 deg) of the fresh gas flow. Increasing window pollution is visible in this latter case, masking velocity information from the IRZ.

In contrast, dramatic differences of the flow behavior between stable and pulsating flame can be detected in the single shot images. Each single shot image in Fig. 6 is represented by velocity magnitude of the two measured components (top) and additionally the axial component (bottom). Consecutive velocity maps of the stable 3 bar flame (higher staging) exhibit the same general flow shape, differing only by minor turbulent structures (two top rows).

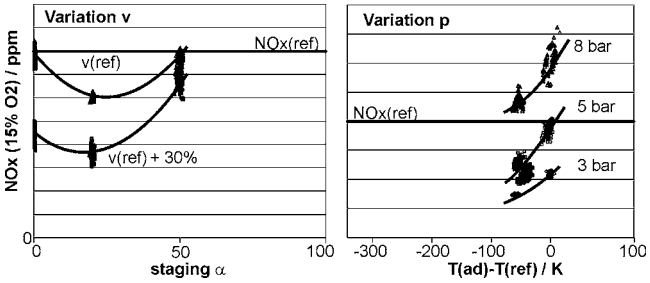


Fig. 3 NO_x value dependency on burner exit velocity (left, different staging, constant *T*(ad)), and pressure (right, constant α).

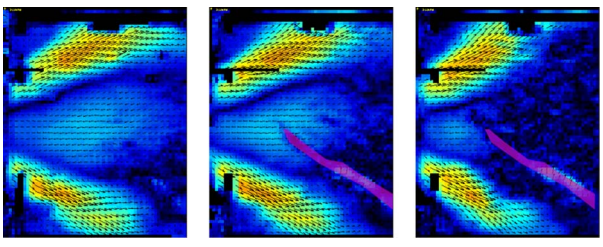


Fig. 5 PIV averages for pulsating flames at low α and 3 bar, 5 bar, and 8 bar (left to right), velocity scale as before; a crack in the window (masked), and increasing window pollution is visible for the 5 bar, and 8 bar flames.

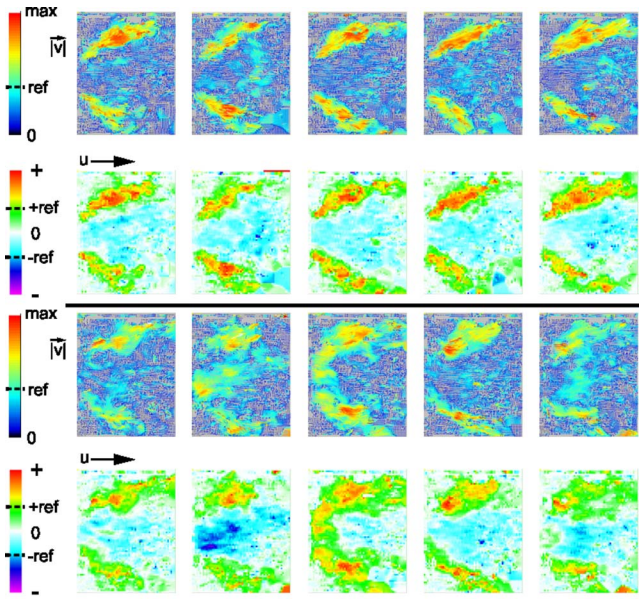


Fig. 6 PIV single shots for a nonpulsating (top, 3 bar, intermediate α) and a pulsating flame (bottom, 3 bar, same T(ad), low α). Each flame snapshot is represented by velocity magnitude (top) and axial component (bottom). In the pulsating case, a selection of strongly different images is shown while the stable flame is visualised as sequence.

The high-velocity regions (red in the top row) remain very similar from image to image. The u component indicates moderate, but temporally very stable negative velocities in the IRZ. The opposite is true for the pulsating case: The color pattern of the velocity magnitude (Fig. 6, third row) is strongly different in the presented images. Velocity vectors in the inner recirculation zone close to the burner exit change direction and magnitude on a large scale and even the shape of the main gas injection changes (Fig. 6, two bottom rows, see especially changing colors in lower row, indicating the axial velocity component u). Large vortices appear in the shear layer, changing the shape of the flow significantly in the axial and radial direction (latter not shown). Together with the OH experiment, this gives evidence that mixing is strongly enhanced due to flame pulsations relative to the pure turbulent mixture. This strong additional mixing mechanism might be responsible for the significant NO_x reduction of flames in the pulsating mode.

Spectroscopic Data. In lean premixed flames, the emission intensity of OH^* radicals, formed in the reaction zone, is often used as indicator of the local heat release. Even though the influence of composition of the gas mixture, preheat temperature, and strain rates is not completely understood [3,17], this emission can be used for qualitative imaging of the heat release in the present case because the mentioned parameters remain similar in the studied flames. In contrast, LIF of OH in the electronic ground state indicates hot exhaust gas behind the flame front. Because of strong laser absorption on the relatively long pathway through the flame, the OH LIF signal was too low for analysis in the lower parts of the combustion chamber. Therefore, only LIF images from the upper half are shown in the following. Note that only the most upstream laser-sheet position could be analyzed. Further downstream, the inner window surfaces suffered too much from direct flame impact preventing proper laser transmission. Nevertheless, the most important regions with respect to OH LIF have been covered by the experiments as will be shown below.

Pronounced pressure pulsations could be detected, beginning with ignition at atmospheric pressure. These were used for triggering phase-resolved measurements. During combustor stabilization (heating up the walls), the pulsation frequency increased. At increased pressure, pulsations were prominent for small staging and not significant for higher staging.

Among the significantly pulsating flames the one at 5 bar, T(ad)=T(ref), α=0, v=v(ref) exhibited the most significant behavior (Fig. 7) and therefore was used for validation of the LES computations. A very pronounced oscillating mode at 245 Hz is visible in the averaged OH^* images (Fig. 7(*a*)). Even though the flame is not perfectly axisymmetric, a deconvolution of these line-of-sight integrated images is possible; this gives a better impression of the processes occurring in the central plane (Fig. 7(*b*)). The deconvoluted images indicate that the heat release is mostly not located near the burner axis but in a conically shaped region, and strongly changes intensity and position. The heat release takes place predominantly close to the shear layer, between the injected fresh gas and the inner recirculation zone, characterized by strong turbulent mixing. The flame position varies between anchored to

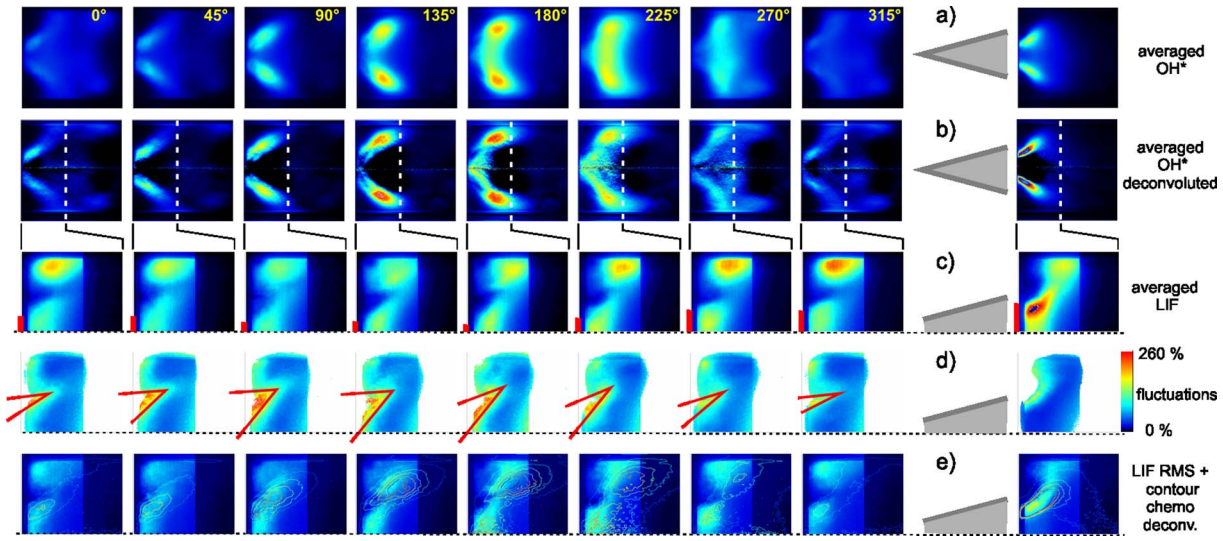


Fig. 7 OH^* and OH LIF measurements of strongly pulsating flame at 5 bar, T(ad)=T(ref), α=0, v=v(ref) at 245 Hz (left) and comparison to the stable flame at intermediate α (right). The approximate burner geometry, relative to the flame shown in the right column, is indicated in one extra column. Note that a different window section is chosen for the tree bottom rows.

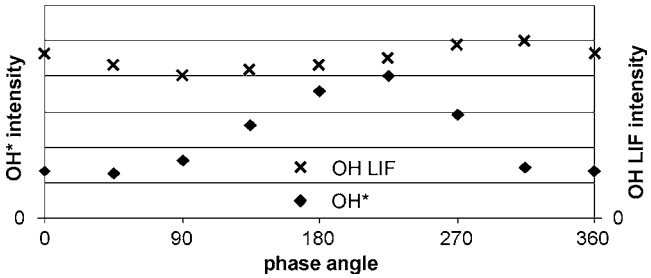


Fig. 8 Full image integrated intensities for the flame shown in Fig. 7. The OH* line represents a full 3D integration, while the OH LIF trend describes only a 2D integrated intensity.

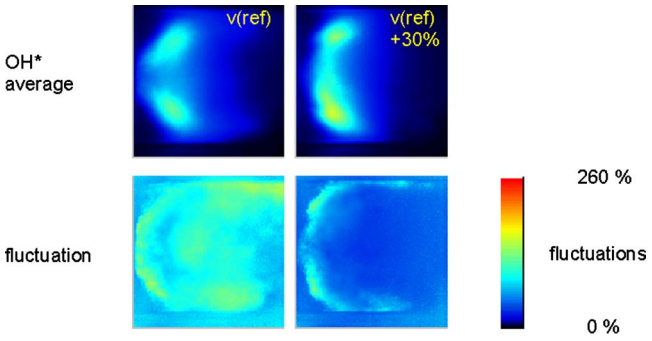


Fig. 10 Detailed view on fluctuation of OH* in phase averaged images on variation of burner exit velocity, left: 5 bar flame from Figs. 7, right: 5 bar flame shown in Fig. 9.

the burner (opening angle of <80 deg at low and high phase angles) and lifted (>95 deg at intermediate phase angles). Moreover, the intensity integrated over the full image shows a significant periodic fluctuation of integral heat release with a maximum around 225 deg (Fig. 8). Compared to the stable flame at intermediate α (right column, in Fig. 7) the spatial region of heat release is much more distributed during the entire pulsation. The corresponding LIF images of OH (hot exhaust gases, (Fig. 7(*c*)) show a change of intensity mainly close to the windows (upper edge of images) and a signal-free region in the position of fresh gas injection that is not visible in the stable case at intermediate α. Visualization of the relative LIF signal fluctuations (in Fig. 7(*d*)) even more clearly shows changes in the region of the cold air injection (low signal, no fluctuation) and shear layer with the inner recirculation zone IRZ (strong turbulence, strong heat release, thus large relative fluctuations). Structural changes of the inner recirculation zone, indicated by gray (red) lines, are visible in the relative fluctuation images and the OH LIF averages. At 90–180 deg, the OH distribution is more lifted, while at 315 deg, it is anchored (vertically wide inner recirculation zone spreading into the burner, gray (red) bar in third row). A spatial correlation of LIF rms and deconvoluted OH* (Fig. 7(*e*)) shows a transition of heat release from the aforementioned shear layer close to the burner (0 deg) towards a position close to the outer recirculation zone distant from the burner (135/180 deg). The position of the OH LIF maxima outside and sometimes upstream from the heat release region indicates the existence of the mentioned ORZ, which seems to participate significantly in proper flame stabilization.

Figure 8 shows that the acoustic pulsation is similarly imaged by the integrated OH* and OH LIF signal with a phase delay in between. A heat-release variation during the oscillation by a factor of 3 induces strong temperature fluctuations correlated with varying OH concentrations. Note that differences in the amplitudes shown in Fig. 8 are induced by the different integration regimes (the OH LIF integration is only two-dimensional). Because the LIF images are dominated by the outer recirculation zone (ORZ), the delay is probably due to the transport time of OH radicals between formation in the shear layer and the position of OH maxima close to the windows in the ORZ.

A very similar behavior was detected for a 3 bar pulsating flame (T(ref), low α, not shown) at the same burner exit velocity. In contrast, starting from the pulsating 5 bar operating point and increasing the burner exit velocity by 30%, the pulsation intensity is acoustically conserved but the optical pulsation behaves significantly different (Fig. 9). In this case, the position of heat release no longer changes. The main fluctuations are in intensity, but not that prominent as above, with the flame lifted at all phase angles. A minor asymmetry is visible in the images, indicating the high sensitivity of quasi-pre-mixed flames to outer parameters; a reason for that behavior was not detected during the experiments. In the LIF images, no significant change is visible neither for position of OH nor for signal intensity. Because of the higher burner exit velocity, the air and fuel inlet seem to behave more stiffly from an acoustical standpoint. The acoustically stiffer behavior at higher burner exit velocity and the correlated weaker breathing character is clearly visible in Fig. 10, showing phase-averaged intensities and fluctuations of the pulsating flames differing only in burner exit velocity. Relative fluctuations are significant only close to the burner for the higher velocity case (shaded (yellow-green)), while at low burner exit velocity the fluctuation level is generally higher and spread over the complete combustion chamber region. Dark (blue) indicates low relative fluctuations.

Besides a detailed study of the pulsations of the downscaled technical burner, the experiments provided insight into the burner behavior as a function of varied operating parameters. Figure 11 (left) illustrates changes of the flame shape with the staging ratio at 5 bar. Heat release is very diffuse and lifted at low α while strongly localized, bright, and anchored to the burner at high α. The already-mentioned separation of inner and outer recirculation zones by the fresh gas injection is prominent in the OH LIF images for low staging only; the single-shot images indicate the presence even at higher α, but due to statistics, this effect is not as pronounced as for α=0, where every laser shot induces a signal-free region in the same position. The trend in the averaged LIF

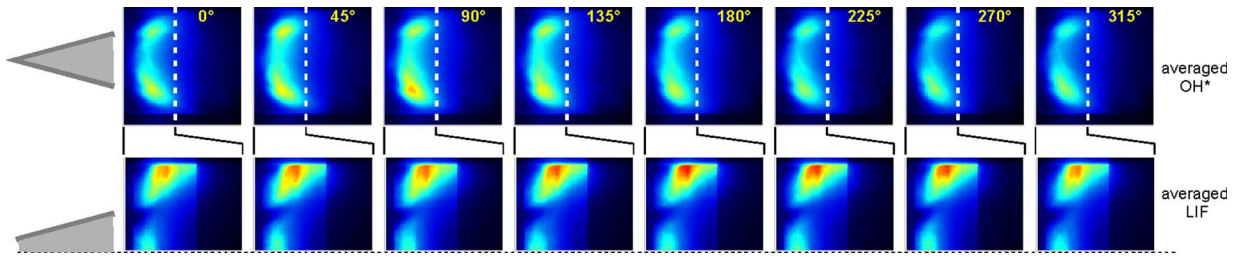


Fig. 9 OH* and OH LIF measurements of an acoustically pulsating flame, parameters as for flame in Fig. 7, but 30% higher burner exit velocity, 300 Hz.

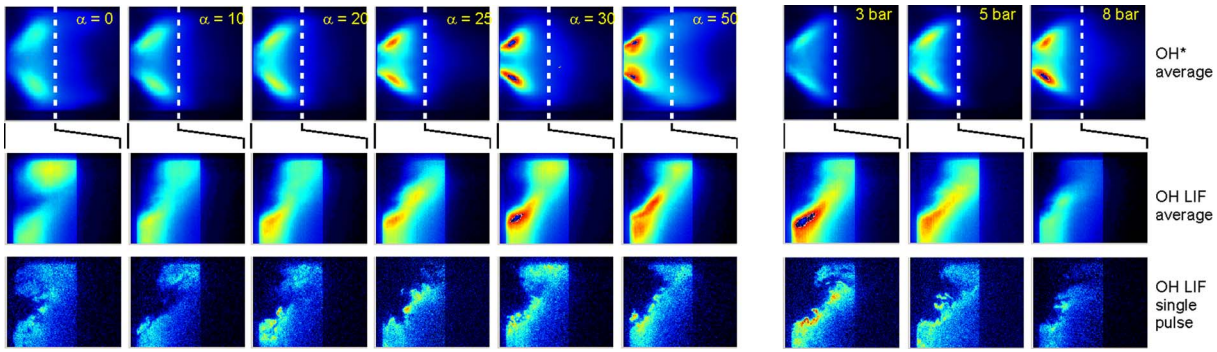


Fig. 11 Left side: OH* and OH LIF trends on variation of staging, 5 bar, *T*(ad) and *v*(exit)=const, left to right: Increased staging. Right side: Trends on variation of pressure at constant T_{ad} and staging.

images is dominated by a displacement of the signal maximum from the pronounced outer recirculation zone at low α toward a position in the inner recirculation zone close to the burner at high α.

An increase of pressure at similar burner exit velocity does not result in a significant change of flame shape. This trend is shown for an intermediate staging ratio in Fig. 11 (right); not shown is the comparison for low α, where the effect looks the same but at different flame shape. The main effect is an increase of heat release at higher pressure and thus higher (fuel and air) mass flows. Initially surprising is the reverse effect in OH LIF, i.e., a decrease of signal with pressure. Probably this effect is caused by superequilibrium OH concentrations at lower pressures. The single-shot images (bottom) indicate high signal intensity on the fresh gas side of the OH clouds. More distant regions show a more homogeneous, diffuse OH LIF signal structure. If pressure increases, then the lifetime of these nonequilibrium OH molecules is strongly reduced due to molecular collisions, leading rapidly to equilibration.

The influence of the burner exit velocity was studied at 5 bar and different operating conditions. Figure 12 (column 1 and 2) shows a comparison for *T*(ref)-50 K and intermediate α, followed by the same operating point at higher temperature and finally at reduced staging in steps down to α=0. In general, the opening angle visible in the OH* images is wider for the higher burner exit velocity. For lower velocity conditions, the flame position is more anchored to the burner, even at low staging. As mentioned above in the context of pulsations, the *T*(ref), α=0, *v*(ref)+30% flame is lifted during all phase angles of the instability, shown here in the context of the phase-averaged images. The LIF images additionally show a more distinct separation of inner and outer recirculation zones (hot exhaust gases, including OH radicals) for the higher velocity flames, while the intensity maxima are shifted toward the combustion chamber windows (moderately for higher α, very pronounced for low α). Two effects come into consideration to explain this behavior: The pure existence/nonexistence of a prominent outer recirculation zone or a movement of the OH concentration maxima out of the measurement plane defined by the laser due to the changed, especially tangential component of the fresh gas velocity. The changed shape of the OH* images (top row) favors the first explanation. Furthermore, a stabilization of the flame in a lifted mode is known to depend on high enough burner panel temperatures [18], as can be realized by hot exhaust gas recirculation in the ORZ. Obviously, the main flow directions of the combustor are modified under these conditions.

Simulation Results. For validation of the numerical model, one significantly pulsating test case was chosen (Fig. 7). In addition to flow rates, pressure and media temperatures, system data of the test rig, such as wall temperatures, exhaust gas temperature, and pressure drop, were helpful for definition of boundary conditions for numerical simulation.

Modeling the test flame with nonreflecting outlet conditions did not result in any combustion instabilities, neither with adiabatic nor with isothermal walls and without accounting for radiative losses in both cases. Nevertheless, the flow field looks relatively similar regarding time averages with or without reflecting outlet. Any of the different simulations: (i) Adiabatic walls, no radiation,

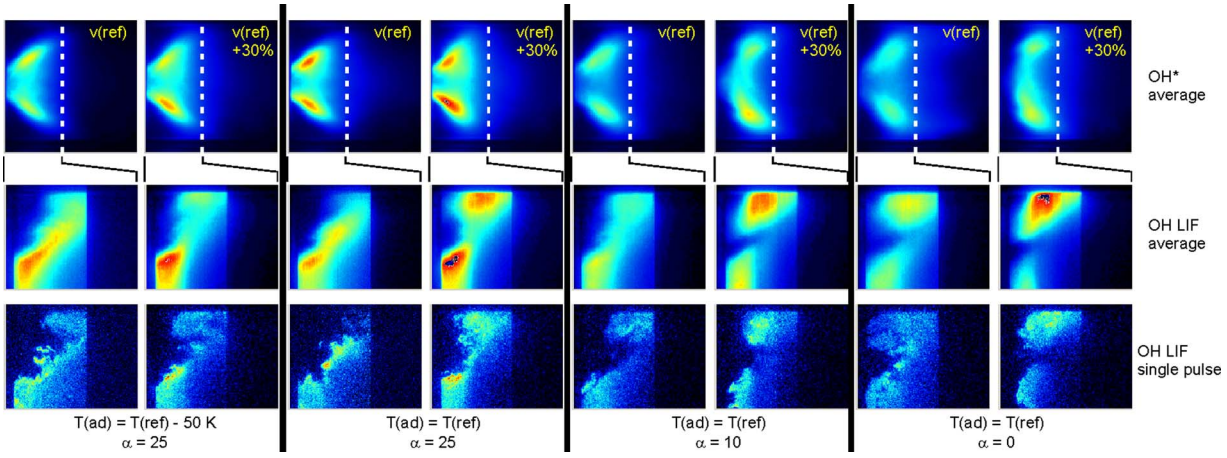


Fig. 12 OH* and OH LIF trends on variation of burner exit velocity, 5 bar. Comparison for intermediate low α, lower flame temperature (left) to α=0, higher temperature (right). Each column contains lower (left) and higher velocity (right) images.

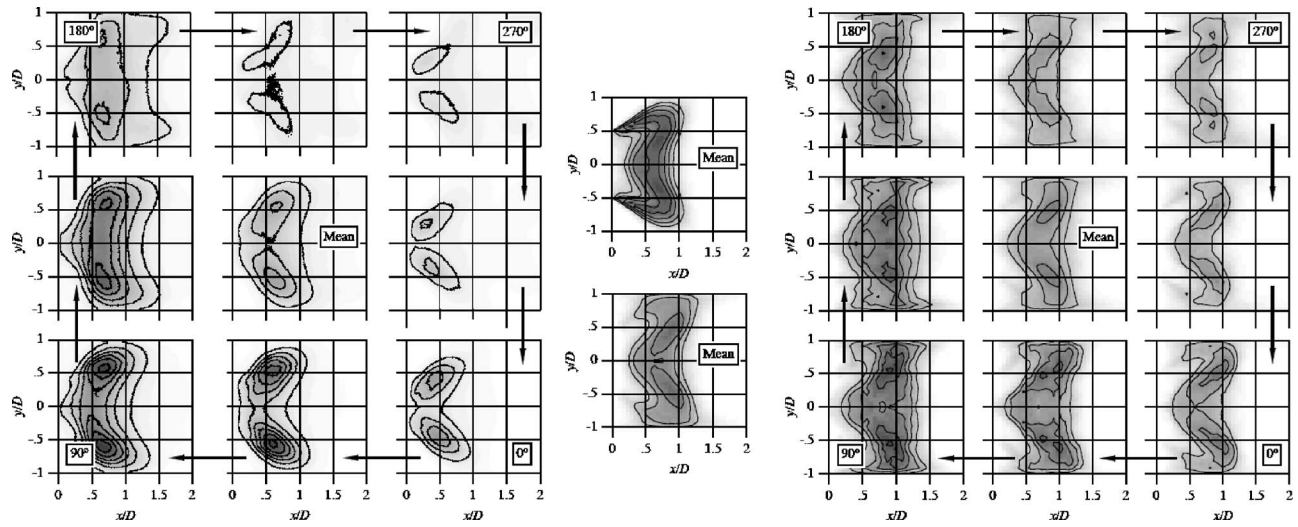


Fig. 13 Experiment (left, flame shown in Fig. 7) and model results (right) match well. If assuming the walls to be adiabatic and without reflective combustion chamber outlet (middle top) or isothermal nonreflective (middle bottom), the agreement is insufficient, yet better for the isothermal case. For all plots, the intensity scale is linear; contour lines indicate same values for all computations.

non reflecting boundaries; (ii) isothermal walls, nonreflecting boundaries; and (iii) isothermal walls, reflecting boundaries, shows good agreement with the experimental data with respect to radial and axial mean velocities. Axial fluctuations match relatively well, and only radial fluctuations in the shear layer are overpredicted.

The plots in Fig. 13 apply a linear scale. Since no absolute values have been measured, the description of the relative behavior of the flame on the instability is important for this comparison of theory and experiment. Comparing the mean line-of-sight integrated heat release (phase averaged for the reflective case) shows clear differences in the calculated flame shapes. Although in the first case (Fig. 13, central column top) significant heat release is predicted on the burner axis; this is less (case 2, central column bottom) or not true (middle image right matrix) for the other two approaches, the latter strongly resembling the experiment (middle image left matrix). Moreover, at a calculated combustion instability of 275 Hz, the phase resolved images show a good agreement with the experimental data with regard to flame structure and intensity fluctuation (compare matrices).

Although the reflecting boundary condition (case 3) is essential for predicting a pulsation at all, as described above, even a pure improvement of the thermal description (case 2) yields a more realistic time integrated flame shape. The calculated heat losses for radiation and convection sum up to 14%; thus, the influence on NO_x formation is much stronger. When assuming adiabatic wall conditions and no radiation, the flame temperature remains at adiabatic values downstream the flame front causing a continuous increase of thermal NO production until the chamber outlet. Inclusion of thermal losses reduces the temperature in the downstream part of the combustion chamber, and thus the fraction of thermal NO remains very low. One order of magnitude of emitted NO_x between cases 1 and 2 clearly indicates that proper thermal balancing should be considered more than an extension of the used basic kinetic scheme. Nevertheless, besides this general finding, the calculations underestimate NO_x production with respect to experimental data. Better results are probably accessible with a somewhat larger kinetic mechanism, but thus increasing computational efforts and, consequently, taking radiative losses into account in the LES simulation as done in the presented simulations.

Conclusions and Outlook

A downscaled technical burner with a combustion chamber allowing for full optical access has been operated successfully up to 8 bar. The general behavior of full-size, full-power combustors is reproduced. Laser diagnostic tools have been applied to this combustor at various operating conditions. Selected fluctuating flames were studied resolving the underlying pulsation.

Evidence of a strong outer recirculation zone transporting hot exhaust gas back to the flame base in pulsating flames is given. One pronounced fluctuating case has been used for validation of complex numerical simulations. Therefore, inclusion of proper thermal balancing radiation and walls, as well as suitable acoustical reflection parameters was shown to be essential. If doing so, the agreement with experiment is satisfactory. The strong influence of proper thermal balancing found in simulations might give an explanation for the relative high LBO temperatures when combining this information with the existence of non-negligible but experimentally required cold components of the combustion chamber. The setup numerical scheme shall be applied to further validation data that are presented in this paper.

Acknowledgment

Experiments and numerical calculations were performed within the EU Project "Demonstration of a low NO_x fuel-staged combustor in a high-efficiency Gas Turbine" (FUELCHIEF, NNE5/382/2001).

References

[1] Lefebvre, A. H., 1999, *Gas Turbine Combustion*, 2nd ed., Taylor & Francis, New York.
[2] McManus, K. R., Poinsot, T., and Candel, S. M., 1993, "A Review of Active Control of Combustion Instabilities," Prog. Energy Combust. Sci., **19**, pp. 1–29.
[3] Schuermans, B., Bellucci, V., Guethe, F., Meili, F., Flohr, P., and Paschereit, C. O., 2004, "A Detailed Analysis of Thermoacoustic Interaction Mechanisms in a Turbulent Premixed Flame," ASME Paper No. GT2004-53831.
[4] Schmitt, P., 2005, "Thermally Accurate LES of the Stability-Emission Performance of Staged Gas-Turbine Combustion," Ph.D. thesis, INP Toulouse.
[5] Doebbeling, K., Eroglu, A., Joos, F., and Hellat, J., 1999, "Novel Technologies for Natural Gas Combustion in Turbine Systems," Eurogas, **99**, pp. 75–81.
[6] Giezendanner-Thoben, R., Meier, U., Meier, W., Heinze, J., and Aigner, M., 2005, "Phase-Locked Two Line OH-PLIF Thermometry in a Pulsating Gas

Turbine Model Combustor at Atmospheric Pressure," Appl. Opt., **44**, pp. 6565–6577.
[7] Raffel, M., Willert, C., and Kompenhans, J., 1998, *Particle Image Velocimetry*, Springer-Verlag, Berlin.
[8] Willert, C., and Jarius, M., 2002, "Planar Flow Field Measurements in Atmospheric and Pressurized Combustion Chambers," Exp. Fluids, **33**, pp. 931–939.
[9] Moureau, V., Lartigue, G., Sommerer, Y., Angelberger, C., Colin, O., and Poinsot, T., 2005, "Numerical Methods for Unsteady Compressible Multi-Component Reacting Flows on Fixed and Moving Grids," J. Comput. Phys., **202**, pp. 710–736.
[10] Colin, O., and Rudgyard, M., 2000, "Development of High-Order Taylor-Galerkin Schemes for Unsteady Calculations," J. Comput. Phys., **162**, pp. 338–371.
[11] Smagorinsky, J., Manabe, S., and Holloway, S., 1963, "General Circulation Experiments With the Primitive Equations," Mon. Weather Rev., **91**, pp. 99–164.
[12] Poinsot, T., and Lele, S., 1992, "Boundary Conditions for Direct Simulations of Compressible Viscous Flows," J. Comput. Phys., **101**, pp. 104–129.
[13] Baum, M., Thevenin, D., and Poinsot, T., 1994, "Accurate Boundary Conditions for Multispecies Reacting Flows," J. Comput. Phys., **116**, pp. 247–261.
[14] Colin, O., Ducros, F., Veynante, D., and Poinsot, T., 2000, "A Thickened Flame Model for Large Eddy Simulations of Turbulent Premixed Combustion," Phys. Fluids, **12**, pp. 1843–1863.
[15] Grötzbach, G., 1987, "Direct Numerical and Larger Eddy Simulation of Turbulent Channel Flows," *Encyclopedia of Fluid Mechanics*, Gulf Publishing Company, Houston, TX, Vol. 6, pp. 1337–1391.
[16] Barlow, R., Karpetis, A., Frank, J., and Chen, J.-Y., 2001, "Scalar Profiles and NO Formation in Laminar Opposed-Flow Partially Premixed Methane/Air Flames," Combust. Flame, **127**, pp. 2102–2118.
[17] Lee, J. G., and Santavicca, D. A., 2003, "Experimental Diagnostics for the Study of Combustion Instabilities in Lean Premixed Combustors," J. Propul. Power, **19**, pp. 735–750.
[18] Aigner, M., DLR, private communication.

Characterization of Fuel Composition Effects in $H_2/CO/CH_4$ Mixtures Upon Lean Blowout

Qingguo Zhang

David R. Noble

Tim Lieuwen

School of Aerospace Engineering,
Georgia Institute of Technology,
Atlanta, GA 30332-0150

This paper describes measurements of the dependence of lean blowout limits upon fuel composition for $H_2/CO/CH_4$ mixtures. Blowout limits were obtained at fixed approach flow velocity, reactant temperature, and combustor pressure at several conditions. Consistent with prior studies, these results indicate that the percentage of H_2 in the fuel dominates the mixture blowout characteristics. That is, flames can be stabilized at lower equivalence ratios, adiabatic flame temperatures, and laminar flame speeds with increasing H_2 percentage. In addition, the blowoff phenomenology qualitatively changes with hydrogen levels in the fuel, being very different for mixtures with H_2 levels above and below about 50%. It is shown that standard well stirred reactor based correlations, based upon a Damköhler number with a diffusivity ratio correction, can capture the effects of fuel composition variability on blowoff limits. [DOI: 10.1115/1.2718566]

Keywords: syngas, lean blowout, hydrogen

Introduction

This paper describes measurements of the dependence of lean blowout limits upon fuel composition. This work is motivated by interest in developing fuel-flexible combustors capable of operating with highly variable and potentially low quality fuels, while producing minimal air pollutants. Such plants would be useful in a range of settings and geographic locations by having capabilities for efficiently and cleanly operating with a range of fuels, such as liquids or synthetic gases derived from other organic sources, e.g., coal, sewage gas, biomass, or landfill gas [1].

Current dry low emissions technology primarily focuses on burning natural gas, a fuel that is mainly composed of methane. It seems clear, however, that natural gas cannot be relied upon as the exclusive source for fueling the clean power plants of the future. Rapidly increasing demand due to new installations has caused substantial price volatility and concerns about future supplies. In addition, interest in utilizing other energy resources, as well as concern about energy security have motivated interest in utilizing coal-derived syngas or fuels from other sources, such as biomass, landfill gas, or process gas. Technologies such as integrated gasification combined cycle (IGCC) plants enable the combustion of coal and other solid or liquid fuels, while still maintaining aggressive emissions targets and high efficiency.

The inherent variability in composition and heating value of these fuels provides one of the largest barriers towards their usage, however. Syngas fuels are typically composed primarily of H_2, CO, and N_2, and may also contain smaller amounts of CH_4, O_2, CO_2, and other higher order hydrocarbons [2]. Depending upon the source and particular processing technique, these fuels can have significant ranges in relative composition of these constituents. To illustrate, Moliere [3] presents a comparison of compositions for 12 different syngas fuels, showing that the volumetric H_2/CO ratio varies from a low of 0.33 to a high of 40, the percentage of diluents gases (e.g., N_2, CO_2, Ar) from 4 to 51%, and the percentage of water from 0 to 40%. The primary constituents of landfill or sewage gas are typically CH_4 and CO_2 [4].

This variability is a significant problem because low emission combustion systems are optimized for operation within a tight fuel specification. Expensive test programs and hardware modifications are generally required if there are significant changes to gas fuel properties. This variability in fuel content, and resultant combustion kinetics, will likely introduce substantial modifications in steady state and dynamic combustor behavior from one syngas to the next.

This paper describes results from an experimental program that is investigating an issue that will be of acute significance toward realization of low emissions, fuel-flexible combustors: lean blowout. Blowout can be expected to be a serious issue, given the substantial variability in chemical kinetic properties of the various fuels encountered in fuel flexible combustors, particularly because achieving low emissions often requires operating very near the blowout limits of the system.

Background

Flame stabilization involves competition between the rates of the chemical reactions and the rates of turbulent diffusion of species and energy. While a significant amount of fundamental understanding of flame propagation and stability characteristics of lean, premixed systems has been gained in conventionally fueled, natural gas-air systems [5], little is known about these issues for alternate gaseous fuels, such as syngas or low BTU fuel mixtures. Furthermore, the majority of the fundamental investigations of the combustion characteristics of these synthetic gases are for nonpremixed flame configurations [6–10].

Consider syngas fuels which are composed primarily of H_2, CO, and N_2. Both H_2 and CO act as fuels when oxidized. Individually, they produce higher adiabatic flame temperatures (at stoichiometric conditions in air) than CH_4, 2383 and 2385 K, as compared to 2220 K. H_2 and CO also have lower flammability limits (ϕ=0.14 and 0.34, respectively) than CH_4 (ϕ=0.46), and higher maximum adiabatic laminar flame speeds (320 and 55 cm/s versus 40 cm/s for methane) [11]. This would suggest

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received August 15, 2006; final manuscript received December 26, 2006. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2005: Land, Sea and Air (GT2005), June 6–9, 2005, Reno, NV, Paper No. GT2005-68907.

that these fuels could have better flame holding characteristics than natural gas. The picture for the pure fuels becomes slightly more complicated at lean mixtures, however, because the flame speed of CO drops below that of CH_4. This comparison of pure fuels does not, however, paint a complete picture. The turbulent propagation and stability properties of premixed H_2/CO mixtures are not well documented, and performance characteristics of these mixtures cannot be simply inferred from knowledge of the global performance of the constituents. To begin, CO and H_2 have *significantly* different transport properties and flame speeds. Next, CO chemistry, which releases slightly more heat than the same amount of H_2 by mass, is highly coupled to H_2 oxidation through the reaction $CO+OH \rightarrow CO_2+H$, which dominates CO ignition (at least) at atmospheric pressure.

The coupling is also important in that, together, CO and H_2 oxidation can cover a wide range of time scales that may bridge the relevant fluid dynamic, turbulent time scales. As shown below, chemical time scales of CO/H_2 mixtures are quite different, and given the large spectrum of turbulent time scales in practical combustors, the combustion processes can cover several orders of magnitude of the Damköhler number. For example, interactions between turbulent mixing processes and, say, H_2 oxidation may occur in the "flamelet" regime, those associated with CO oxidation in a "torn flamelets" regime, and OH recombination (whose chemistry plays a key role in CO oxidation) in the distributed reaction regimes. Furthermore, the oxidation and, hence, rates of heat release of CO and H_2 display different or opposite sensitivities to mean pressure or flame strain/stretch imposed by the turbulent flow field.

In addition, syngas fueled plants sometimes co-fire with a certain fraction of natural gas, so the additional interactions associated with H_2/CH_4 and CO/CH_4 flames also need to be considered. Several recent studies have focused on H_2/CH_4 flames [12,13] and have shown that addition of H_2 enhances the mixture's resistance to extinction or blowoff. For example, fundamental studies in opposed flow burner geometries show that the extinction strain rate of methane flames is doubled with the addition of 10% H_2 [14,15]. The acceleration of the CH_4 reaction rate by the radical pool created by the early breakdown of H_2 has been suggested as a key mechanism for this behavior.

CO/CH_4 flames also have interesting dynamics because of their coupled chemistry. The CO burning rate is highly dependent on the reaction $CO+OH \rightarrow CO_2+H$. The addition of CH_4 increases the radicals for this reaction. In addition, CO is also believed to be a significant intermediate during the low temperature reaction path of CH_4 [16].

The above discussion clearly indicates the need for more extensive systematic studies of the blowout characteristics of premixed flames using the fuels that will be encountered in fuel-flexible combustors. Several studies have been initiated relatively recently to investigate the characteristics of premixed, hydrogen-enriched methane fuels [12,13,17]. Additional studies are needed, however, to broaden the scope of fuels of interest.

Having briefly considered the kinetic characteristics of these fuels, we turn our attention next to issues associated with blowoff. Developing physics-based correlations of blowout behavior is complicated by the lack of understanding of the flame characteristics at the stabilization point. Currently, there is disagreement on whether premixed flames in high turbulent intensity gas turbine environments have flamelet, "thickened" flamelet, or well stirred reactor (WSR)-like properties.

Methods for developing blowout correlations in the latter case (i.e., using WSR scaling ideas) have been studied extensively. Several different theories or physical considerations have been used in past blowout correlation studies, such as those of Zukoski and Marble [18], Spalding [19], or Longwell [20]. As noted by Glassman [11], however, they lead to essentially the same form of correlation. These correlations generally involve relating the blowoff limits to a ratio of a chemical kinetic time and residence time, τ_{chem}/τ_{res}. In well stirred reactor theory, this ratio is often referred to as a combustor loading parameter. It is possible that the recirculation regions that stabilize many high intensity flames, which may have flamelet properties at most other points along the flame, have distributed reactor-like properties; hence, the success in WSR models in correlating blowout behavior.

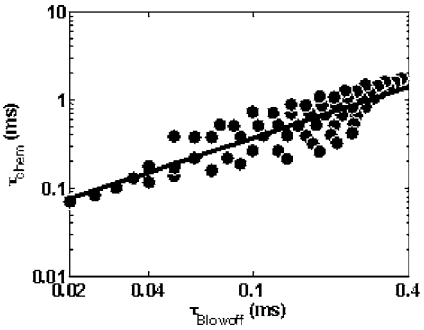


Fig. 1 Relationship between chemical time calculated using Eq. (1) and blowout residence time for $\phi=0.6H_2/CO/CH_4$ mixtures. Results obtained using CHEMKIN with the GRI 3.0 mechanism.

When applied to blowoff limits of premixed flames, this chemical time can be estimated as

$$\tau_{chem} = \alpha/S_L^2 \tag{1}$$

where S_L and α denote the laminar flame speed and thermal diffusivity, respectively [21,22]. Alternative methods of estimating a global chemical time are also possible, but generally lead to results qualitatively similar to Eq. (1). For example, Fig. 1 compares the blowoff residence time, $\tau_{blowoff}$, of a well stirred reactor model to the chemical time from Eq. (1) for several $H_2/CO/CH_4$ mixtures. The two time scales are closely related, with a best fit given by $\tau_{chem}=3.4^*\tau_{blowoff}$ (denoted by the solid line) except for cases with greater than 95% CO (not shown). In this paper, we will use $\tau_{blowoff}$ in estimating chemical times, for the pragmatic reason that they are much simpler and easier to calculate for lean flames. Determination of τ_{chem} is hampered by problems with obtaining converged solutions for S_L at the very low equivalence ratios at which the high hydrogen mixtures blow off.

Returning to WSR based methods for scaling blowoff limits, the residence time scales as d/U_{ref}, where d and U_{ref} denote a characteristic length scale (e.g., a recirculation zone length) and velocity scale, respectively. If this reactor based theory is correct, then blowoff limits should scale with the loading parameter or Damköhler number,

$$\mathrm{Da} = \frac{\tau_{res}}{\tau_{Blowoff}} = \frac{d}{U_{ref}\tau_{Blowoff}} \tag{2}$$

This Damköhler number correlation is equivalent to Peclet number correlations [23] if $\tau_{blowoff}$ is replaced by τ_{chem}. Defining the Peclet numbers based upon flame and flow velocity, $\mathrm{Pe}_u = U_{ref}d/\alpha$ and $\mathrm{Pe}_{SL}=S_Ld/\alpha$ note that

$$\mathrm{Da} = \frac{\mathrm{Pe}_{SL}^2}{\mathrm{Pe}_u} \tag{3}$$

These purely combustion considerations are incomplete without consideration of the corresponding fluid mechanics, however. For example, note that U_{ref} does not need to directly scale with the approach flow velocity, U_0, due to the acceleration of the burned gas [21]. Since the burned gas velocity scale is given by $U_b = (T_b/T_0)U_0$, then $U_{ref}=f(U_0, T_b/T_0)$. Similar considerations apply for the recirculation zone scale, d. For this reason, prior workers have often had to measure the recirculation zone length in bluff body flames in order to use Eq. (2), e.g., see Ref. [18].

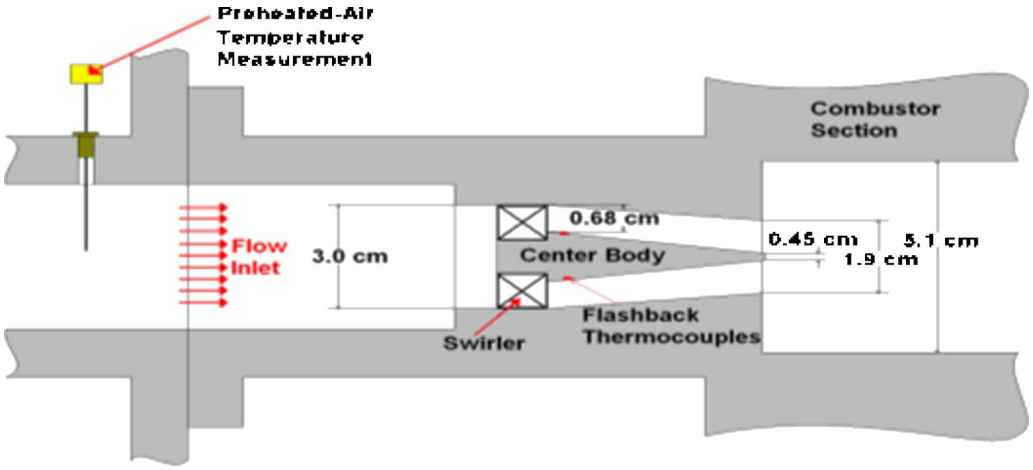


Fig. 2 Cross section of premixer assembly

The work of Hoffman et al. [24] is of special interest, as it found good success with the Peclet number correlation of Eq. (3) to capture the dependence of blowoff limits in swirling, premixed flames upon combustor diameter, flow velocity, and swirl number. They used the azimuthal velocity component, U_θ, as the reference velocity, $U_{\text{ref}}=U_\theta$, and combustor diameter, D, as the characteristic length, $D=d$.

Significantly less attention has been given to correlating premixed flame blowout limits assuming flamelet-like combustion properties, where the stabilization mechanism is related to front propagation, rather than reactor extinction. In this case, a flame would blow off when the turbulent flame speed is everywhere less than the flow velocity, $S_T < U_{\text{ref}}$, where S_T denotes the turbulent flame speed. If this propagation mechanism is controlling, then blowoff limits should scale with the parameter,

$$L_2 = \frac{S_T}{U_{\text{ref}}} \tag{4}$$

Evaluating the turbulent flame speed introduces additional complications for mixtures with widely varying compositions. Correlations of the turbulent flame speed of the form, $S_T = S_L \cdot f(u', \text{geometry})$, where u' denotes turbulent intensity, have been used successfully in many prior studies across limited fuel ranges. However, recent studies across broader ranges of fuels clearly indicate the limitations of the above correlation; other fuel properties are also very important. For example, Kido et al. [25] measured the dependence of the turbulent flame speed for a variety of H_2, CH_4, and C_3H_8 mixtures. These mixtures were carefully chosen to have identical laminar flame speeds, as indicated by the S_T curves converging at $u'=0$. Interestingly, however, the curves widely diverge as u' increases from zero, e.g., the lean hydrogen mixture has a turbulent flame speed almost ten times higher than the lean propane mixture. We have not pursued correlations of this form due to lack of data on S_T for the fuel blends of interest.

Instrumentation, Experimental Facility, and Experimental Procedures

Measurements were obtained in a lean, premixed gas turbine combustor simulator, which has also been previously described in Ref. [26]. The facility consists of inlet/premixer, combustor, and exhaust sections. High-pressure natural gas and air are supplied from building facilities. The air can be preheated up to 700 K. The hydrogen and carbon monoxide are supplied from bottles. The air and fuel flow rates are measured with a critical orifice and mass flow controllers (MFCs), respectively. Both the orifice and MFCs were calibrated using the specific gas with which they were to meter. This is necessary for H_2 in particular, as the manufacturer supplied corrections that relate the flow of some other gas to the H_2 flow rate were found to be very inaccurate. The resultant uncertainty in the flow rate measurements is 2% of full scale and in blowoff equivalence ratios is 0.01–0.02 for most of the cases. The largest uncertainty in ϕ of 0.03 occurs in the 39 m/s case, with pure CH_4. In order to ensure that acoustic oscillations did not affect the fuel/air mixing processes, the fuel and air are mixed upstream of a second choke point. Thus, the equivalence ratio of the reactive mixture entering the flame is constant. The temperature of the reactants was measured with a thermocouple located just upstream of the swirler.

The fuel-air mixture entered the circular 4.75 cm diameter, 60 cm long inlet section and passed through a nozzle and swirler section prior to entering the combustor, see Fig. 2. The nozzle outer body slightly constricts along the axial flow direction. However, the overall flow area remains constant at 10.8 cm^2, as the center body diameter also decreases in the axial flow direction. This nozzle is fully modular as the center body and swirler can be easily removed and replaced; these tests were performed with a single 12 vane, 35 deg swirler. A thermocouple is imbedded in the center body for flashback detection.

The nozzle terminates into the 5×5 cm square combustor. The square part of the combustor is 51 cm long and optically accessible. It then transitions into a circular 7.6 cm diameter, 195 cm long exhaust section. The exhaust sections are water cooled. The flow leaves the setup through an exhaust nozzle with an adjustable bypass valve. This adjustable bypass valve is controlled in order to maintain the combustor pressure at some prescribed value.

The basic test sequence is to operate at uniformly spaced fuel compositions in $H_2/CO/CH_4$ space, such as is depicted in Fig. 3. At each fuel composition, the mixture equivalence ratio is adjusted (at constant premixer velocity) until the mixture blows off.

The flame was viewed with a standard and UV sensitive (down to 270 nm) video camera, each connected to a monitor in the control room. The second UV sensitive camera was needed at the high pressure condition, where the high H_2 flames, even when well stabilized, were often not visible.

Obtaining these data was complicated by the need to keep the approach flow velocity, combustor pressure, and mixture temperature constant across the range of fuel compositions. As such, approaching blow off limits with a certain fuel composition required simultaneously adjusting the air and three fuel flow rates in order to keep constant approach flow velocity. In addition, due to variations in mixture burned gas temperature, maintaining a constant combustor pressure required simultaneous adjustment of the back pressure valve. Finally, variations in molar volume of the fuel necessitated adjusting the air temperature in order to maintain a

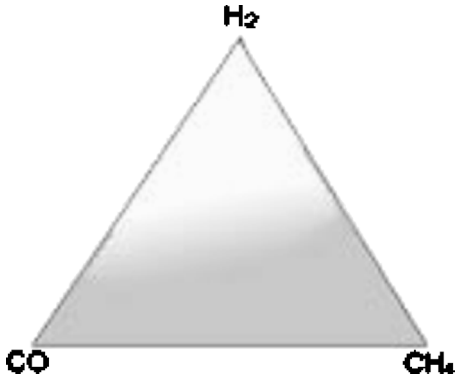


Fig. 3 Composition map describing regions where sharply defined blowoff event occurs (gray) and blowoff preceded by significant flame liftoff (white)

constant reactant temperature. For the data shown in the Results, the approach flow velocity, pressure, and temperature remains constant to within 2%, 5%, and 15 K of their quoted values.

Applying a uniform definition of blowoff is complicated by the fact that the manner in which the flame blew off varied with composition. In many cases, the blowoff event occurred abruptly with a small change in fuel composition, although sometimes preceded by slight liftoff of the flame from the burner. Defining the blowoff point was unambiguous in these instances; moreover, the point of blowoff and flame liftoff was nearly identical. This was the case for mixtures with less than approximately 50% H_2 by volume. However, for high H_2 mixtures, the blowoff and liftoff events were quite distinct, see Fig. 3. Usually, the flame became visibly weaker, lifted off from the holder, and moved progressively downstream with decreases in equivalence ratio before blowing off for good. As such, blowoff is defined here as the point where the flame is no longer visible in the 10.2 cm long optically accessible section of the combustor. Undoubtedly, this variation of liftoff/blowoff characteristics with fuel composition is responsible for some of the scatter in the experimental data. This point should be kept in mind when comparing 0–50% H_2 and 50–100% H_2 containing fuels.

Analysis Approach

This section describes the methods used to postprocess the data and correlate blowout limits. Adiabatic flame temperatures were calculated for a given mixture using standard methods. Laminar flame speeds were calculated with the PREMIX application in CHEMKIN, using the GRI3.0 mechanism. While this mechanism was primarily optimized for methane/air mixtures, good comparisons between its results and measurements have been obtained for a range of H_2/CO mixtures as well [27]. Chemical times, determined with Eq. (1), require the thermal conductivity of the reactive mixture, determined using transport properties from TRAN in CHEMKIN and the equation [28]

$$\lambda = \frac{1}{2}\left(\sum_{k=1}^{K} X_k \lambda_k + \frac{1}{\sum_{k=1}^{K} X_k/\lambda_k}\right) \tag{5}$$

Diffusivity coefficients of a given species in the mixture were determined from

$$D_{km} = \frac{\sum_{j\neq k}^{K} X_j W_j}{\bar{W}\sum_{j\neq k}^{K} X_j/D_{jk}} \tag{6}$$

where W and X denotes the molecule weight and mole fraction, respecitvely. The "effective" diffusivity of the fuel mixture was defined as

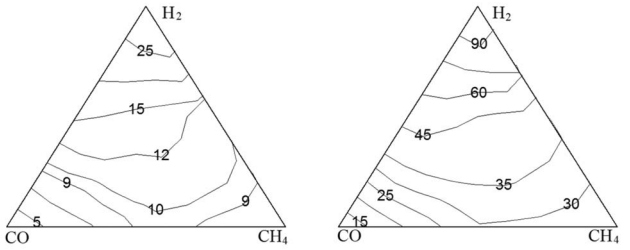


Fig. 4 Dependence of flame speed (cm/s) upon fuel composition at fixed 1500 K (left) and 2000 K (right) adiabatic flame temperatures with 300 K reactant temperature

$$D_F = \sum_{k=\text{fuel}} A_k D_{km} \tag{7}$$

where A_k is the fraction of heat release due to fuel k relative to that of the entire mixture.

In order for the reader to obtain some familiarity with the properties of H_2/CO/CH_4 mixtures, a number of results are included below. Consider first the flame speed, S_L. Figure 4 illustrates the dependence of the flame speed upon fuel composition at two fixed adiabatic flame temperatures, 1500 and 2000 K. The fuel composition by volume is given by the location within the triangle, where the three vertices denote pure CO, H_2, or CH_4. At each point, the mixture equivalence ratio is adjusted such that the mixture has the given flame temperature. As expected, the high H_2 mixtures have the highest flame speeds. Note also the slightly higher flame speeds of the high CH_4 mixtures relative to those of CO mixtures.

An alternative way to view these results is to plot adiabatic flame temperature at a fixed flame speed. This is done in Fig. 5 for $S_L = 10$ and 20 cm/s. Note the progression in flame temperatures from CO and H_2 mixtures being the highest and lowest, respectively.

Figure 6 plots the dependence of the chemical time, $\tau_{\text{chem}} = \alpha/S_L^2$, upon fuel composition at a fixed flame temperature of 1500 K. Note the order of magnitude variation in chemical time from the fast H_2 mixtures to the slow CO mixtures.

Results and Discussion

This section presents blowout results at two constant nozzle exit velocities: 59 and 39 m/s. This corresponds to combustor velocities (cold flow) of 6.0 and 4.0 m/s. Tests were performed at two pressure/temperature conditions: combustor pressure of 1.7 atm and 300 K reactants, and combustor pressure of 4.4 atm and 460 K reactants. The mean equivalence ratios ranged from roughly 0.15 to 0.60.

In order to facilitate presentation of results, we represent the mixture composition of H_2/CO/CH_4 by its color. Primary colors

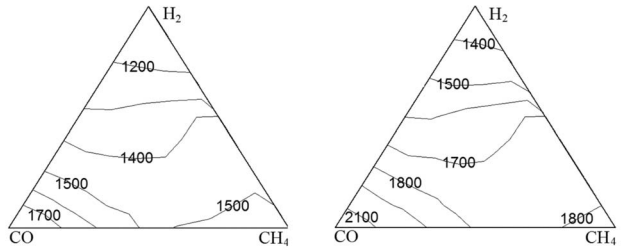


Fig. 5 Dependence of adiabatic flame temperature (K) upon fuel composition at fixed laminar flame speed 10 cm/s (left) and 20 cm/s (right) with 300 K reactant temperature

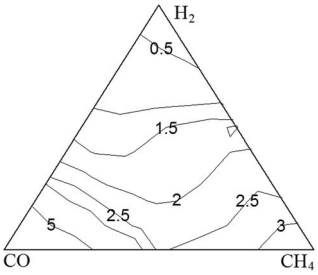


Fig. 6 Dependence of chemical time (ms) upon fuel composition at fixed adiabatic flame temperature, 1500 K

at the three vertices are used to represent each fuel constituent, where red, yellow, and blue denote H_2, CO, and CH_4, respectively. This is illustrated in Fig. 7.

These blowout limits were correlated with a variety of parameters. As noted in previous studies [13], the presence of H_2 has a strong impact on blowout limits of either H_2/CH_4 or H_2/CO flames. Figure 8, which plots the dependence of the blowoff equivalence ratio upon the mole fraction of H_2 in the fuel, shows that H_2 strongly affects the lean blowout limits of syngas mixtures ($H_2/CO/CH_4$), in spite of the complicated coupling chemical mechanisms among these species. It shows the well known result that, in general, mixtures can be stabilized with lower equivalence ratios as the H_2 concentration increases. However, note that the

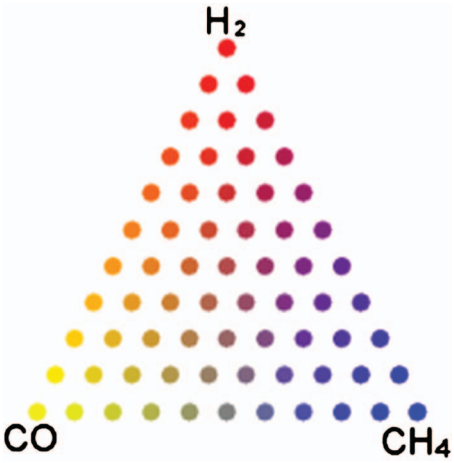


Fig. 7 Primary color mixing scheme used to denote fuel blend composition

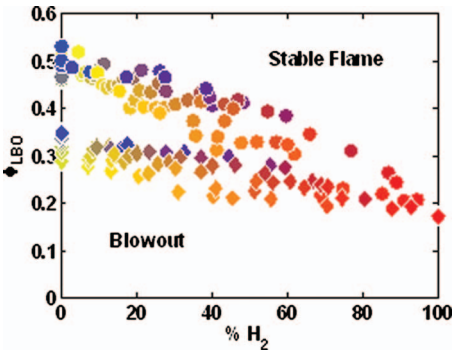


Fig. 8 Dependence of LBO equivalence ratio upon H_2 mole fraction at premixer flow velocities of 59 m/s at 300 K reactants temperature and 1.7 atm combustor pressure (circle) and 458 K and 4.4 atm (diamond)

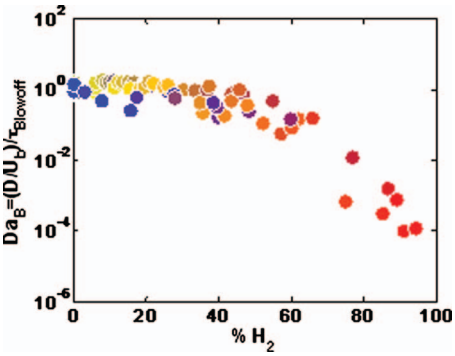


Fig. 9 Damköhler numbers of mixtures at constant premixer flow speed of 59 m/s at 300 K reactants temperature and 1.7 atm combustor pressure

addition of small amounts of H_2 has a small impact upon blowoff limits and that the sensitivity of the blowoff equivalence ratio to H_2 level variations remains essentially constant across the entire range of H_2 levels. In other words, no discontinuous or steep drop in blowoff equivalence ratio was observed with small amounts of H_2 addition.

At a specific H_2 mole fraction there is scatter in the data within a relatively narrow band due to variations in the CO/CH_4 ratio—a higher ratio of CO/CH_4 resulting in blowout at lower equivalence ratios.

The adiabatic flame temperature of the mixture at blowout correspondingly decreases as the percentage of H_2 increases. Although not shown, there is a nearly linear relationship between the percentage of H_2 and adiabatic flame temperature at lean blowout. For example, the adiabatic flame temperature linearly decreased from 1400 K to 1000 K for the high reactant temperature case shown in Fig. 8. This may be partly explained by the chemical kinetics. The chain-branching reaction $H+O_2 \Leftrightarrow O+OH$, which is very sensitive to flame temperature, plays an important role in flames with H_2 addition [15]. For a given fuel composition, when the equivalence ratio decreases to some level where either the flame zone temperature or the concentration of radicals is too low to maintain this chain-branching reaction, local extinction occurs. However, higher H_2 fuels may provide more radicals for this chain-branching reaction, so that flames which have a higher percentage of H_2 could stabilize at a lower adiabatic flame temperature.

Similarly good correlations between the laminar flame speed, Lewis number, and a number of other combustion parameters at blowoff upon H_2 level were observed. This brings us to an important point that must be recognized in extracting an understanding of the blowoff physics from these correlations. First, blowoff limits are clearly a strong function of H_2 levels. Second, many other parameters, such as diffusivities, flame temperature, etc. are also strong functions of H_2 level. As such, it is important to not draw conclusions about blowoff physics only because one can correlate results with parameters that are simply functions of the H_2 percentage. For example, a very nice correlation between T_{ad} versus *Le* at blowout exists, because both of them are functions of percentage of H_2. In other words, regardless of whether the mixture *Le* is a physically meaningful parameter, a good correlation will still be observed. In some sense, this is analogous to correlating T_{ad} versus 2^*T_{ad} at blowoff—obviously, a perfect correlation is observed, regardless of whether this is a physically significant parameter.

Damköhler (Da) number correlations were found to correlate the data over all flow velocities, pressures and temperatures for all mixtures with H_2 levels below 50%, as shown in Fig. 9. In the plot, the reference length scale, *D*, was the combustor width, 5.1 cm. Damköhler numbers were evaluated using both the un-

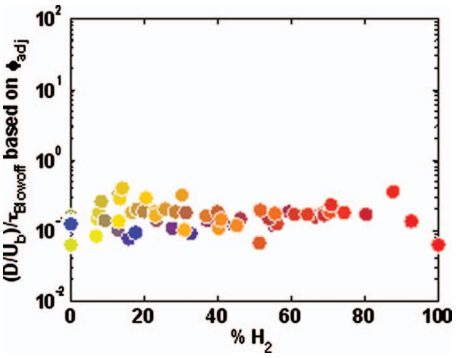


Fig. 10 Damköhler numbers of mixtures based on adjusted equivalence ratio at premixer flow velocities of 59 m/s at 458 K reactants temperature and 4.4 atm combustor pressure

burned, Da_U, and burned flow speed, Da_B, as reference velocity scales. Utilizing the burned gas speed resulted in slightly better ability to correlate the data, as reflected in slightly lower errors (about 10%) in predicted blowoff equivalence ratio, $\delta\phi_{rms}$ (described below), and so is used for these results. Figure 9 shows that blowout occurs at a nearly constant Da for these composition values (although Fig. 9 was plotted in the *logarithm* scale, $\tau_{Blowoff}$ is an exponential function of equivalence ratio). At the same time, Fig. 9 also shows that a constant Da correlation is inadequate for describing blowout limits of higher H_2 level mixtures. It is possible that this is simply a reflection of the fact that the blowout process changes with H_2 levels and that our "blowoff definition" is not the most physically meaningful, see discussion of Fig. 3. For example, perhaps identifying the point where the flame first lifted off the flame holder would have been more useful.

A second possibility for this change in blowoff Da value shown in Fig. 9 may be due to preferential diffusion effects, a consideration that has also been used to scale changes in turbulent flame speed of mixtures whose constituents have significant variations in diffusivity. One approach for incorporating these effects is to note that the local equivalence ratio changes along the wrinkled flame, being both higher and lower than the average at different spatial locations. Kido and co-workers[29] suggested correlating mixture turbulent flame speeds by utilizing mixture properties at an adjusted equivalence ratio, ϕ_{adj}, i.e., not at the average equivalence ratio, ϕ_{ave}, but the average value plus some $\Delta\phi$. As such, mixture properties are correlated at $\phi_{adj}=\phi_{ave}+\Delta\phi$. They suggest the following relation for $\Delta\phi$, based upon empirical fits of their data:

$$\Delta\phi = C\ln(D_F/D_{OX}) \tag{8}$$

where D_F and D_{OX} denote the mass diffusivity of fuel (see Eq. (7)) and oxygen, respectively, and C is an empirical constant whose value they suggest as 0.3.

We found that utilizing a value of C close to 0.1 gives a nearly constant blowoff Damköhler number for all of our data sets, Fig. 10. This plot shows that blowoff occurs at a nearly constant value of local Damköhler number, where the $\tau_{Blowoff}$ is estimated at the equivalence ratio $\phi_{adj}=\phi_{ave}+\Delta\phi$. In this case, the best value for $C=0.07$ and the average value over all the test points of $\overline{Da_B}$ at blowoff is 0.2. Table 1 summarizes the results from the other two tests by presenting the best fit value for C for each individual data set and the corresponding $\overline{Da_B}$ value. It can be seen that $\overline{Da_B}$ does vary with each data set, but is always an O(1) quantity.

In order to quantify the scatter in the correlations shown in the table and the capability for actually inverting the above procedure to be used as a blowoff prediction methodology, we employed the following procedure. Assume that the equivalence ratio at blowoff is now the unknown and must be predicted, $\phi_{LBO,pred}$. Assume also that the Damköhler number at blowoff is known and equal to the value $\overline{Da_B}$ compiled in the table. We then solve the following implicit equation for $\phi_{LBO,pred}$ for each fuel composition:

$$\overline{Da_B} = \frac{d}{U_b\tau_{Blowoff}[\phi_{LBO,pred} + C\ln(D_F/D_{ox})]} \tag{9}$$

This procedure is repeated for each fuel composition. In general, $\phi_{LBO,meas}$ and $\phi_{LBO,pred}$ are not identical, and so the root mean square (rms) of their difference over all the data points is referred to as $\delta\phi_{rms}$. The table also compiles the values of $\delta\phi_{rms}$. As can be seen, assuming a constant Damköhler number at blowoff results in a capability for predicting the equivalence ratio value to within about 0.02–0.04.

Moreover, due to the exponential dependence of $\tau_{Blowoff}$ upon equivalence ratio, varying the precise value of $\overline{Da_B}$ or C does not substantially impact the errors in $\phi_{LBO,pred}$. For example, in the first case above, assuming blowoff occurs at constant values of Da=1.0 or 3, instead of the best fit value of 2.1, results in $\delta\phi_{rms}=0.045$ and 0.043, respectively.

Moreover, both $T_{in}=300$ K, $P=1.7$ atm data sets can be reasonably collapsed with a single $\Delta\phi$ equation or C value. To illustrate, Fig. 11 compares the predicted and actual blowoff equivalence ratios for all low temperature data taken in this study, assuming $C=0.1$ and $\overline{Da_B}=1.7$. It can be seen that the error in $\phi_{LBO,pred}$ is generally less than 0.05, and $\delta\phi_{rms}=0.045$. Moreover, the highest errors are encountered with the very high CO mixtures, which may simply be a manifestation of the sensitivity of high CO mixtures to ambient humidity levels and other factors influencing H levels. If the $P=4.4$ atm, $T=460$ K data set were also plotted, they would also cluster along a line, but with a systematic difference from the grouping in this graph., In other words, the optimum model constants (particularly the C value) vary with operating conditions.

There are a variety of reasons that the remaining scatter could be present, such as inherent noise in the blowoff equivalence ratio.

Table 1 Summary of optimum model constants for correlating blowoff data and resulting scatter in fitted data

	Best fit C value		
Test group	C	Da_B at ϕ_{adj}	$\delta\phi_{rms}$
$T=300$ K, $P=1.7$ atm, $U_0=59$ m/s	0.1	2.1	0.04
$T=300$ K, $P=1.7$ atm, $U_0=39$ m/s	0.08	1.1	0.03
$T=458$ K, $P=4.4$ atm, $U_0=59$ m/s	0.07	0.2	0.02

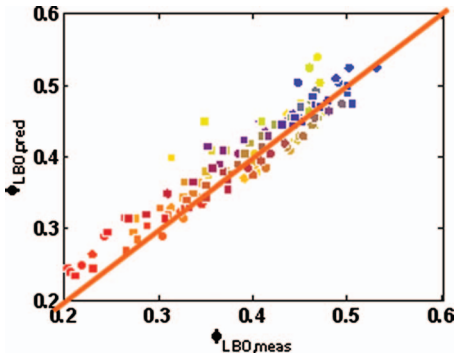


Fig. 11 Comparison of predicted and measured blowoff equivalence ratio for all $T=300$ K, $p=1.7$ atm data. circle: $U_0=59$ m/s, square: $U_0=39$ m/s

In addition, other more subtle factors, such as reference length and reference flow velocity could easily change somewhat with approach flow velocity.

We should emphasize that the C value in the $\Delta\phi$ calculation was chosen *empirically* to give the best fit. Although the Da mechanism, considering preferential diffusion effects, could correlate and predict the lean blowout limits very well, the real physical meaning behind these correlations are still uncertain and require further study.

Concluding Remarks

Although a variety of correlations of varying complexity were assessed, many of which were not reported here, the percentage of H_2 in the fuel was found (for a given set of operating conditions) to be a very powerful parameter for correlating blowoff data over a wide range of fuel compositions. A Damköhler number correlation was found to also correlate the data, even across the range of fuel compositions. The near unity value of this Damköhler number (at least for low H_2 levels), as well as the ability to capture a range of fuel compositions (for $<50\%$ H_2), flow velocities, and ambient conditions suggests that classical well stirred reactor approaches can capture the key blowoff physics. Moreover, the ability of the $\Delta\phi$ correction, motivated by prior studies for correlating turbulent flame speeds, to capture the entire range of H_2 levels is also suggestive of the need to account for differential diffusive corrections when scaling data across broad ranges of H_2 levels. However, we should emphasize that the latter $\Delta\phi$ correction may not necessarily reflect the underlying physics, but simply be another manifestation of the fact that the blowoff limits are a strong function of H_2 levels (note that $\Delta\phi$ is closely correlated with the percentage of H_2). Moreover, comparison of these data at low and high H_2 levels should be done with some caution, given the necessary, but somewhat arbitrary nature of our definition of the blowoff equivalence ratio. In particular, the definition of the blowoff equivalence ratio at the very high H_2 levels, while repeatably given our methodology, could certainly also be done differently, with consequently different values for ϕ_{LBO} at high hydrogen levels.

Acknowledgment

This publication was prepared with the support of the U.S. Department of Energy, Office of Fossil Energy, National Energy Technology Laboratory, under Contract No. 03-01-SR111 (Dr. Richard Wenglarz, contract monitor). Any opinions, findings, conclusions, or recommendations included herein are those of the authors and do not necessarily reflect the views of the DOE.

References

[1] Richards, G. A., McMillian, M. M., Gemmen, R. S., Rogers, W. A., and Cully, S. R., 2001, "Issues for Low-Emission, Fuel-Flexible Power System," Prog. Energy Combust. Sci., **27**, pp. 141–169.
[2] Klimstra, J., 1986, "Interchangeability of Gaseous Fuels—The Importance of the Wobbe Index," SAE paper No. 861578.
[3] Moliere, M., 2002, "Benefiting from the Wide Fuel Capability of Gas Turbines: A Review of Application Opportunities," ASME paper No. GT-2002-30017.
[4] King, S., 1992, "The Impact of Natural Gas Composition on Fuel Metering and Engine Operating Characteristics," SAE paper No. 920593.
[5] Durbin, M., and Ballal, D., 1996, "Studies of Lean Blowout in a Step Swirl Combustor," J. Eng. Gas Turbines Power, **118**, pp. 1–8.
[6] Barlow, R. S., Fiechtner, G. J., Carter, C. D., and Chen, J. Y., 2000, "Experiments on the Scalar Structure of Turbulent $CO/H_2/N_2$ Jet Flames," Combust. Flame, **120**, pp. 549–569.
[7] Correa, S. M., and Gulati, A., 1988, "Nonpremixed Turbulent CO/H_2 Flames at Local Extinction Conditions," Sym. (Int.) Combust., [Proc.], **22**, pp. 599–606.
[8] Drake, M. C., 1986, "Stretched Laminar Flamelet Analysis of Turbulent H_2 and $CO/H_2/N_2$ Diffusion Flames," Sym. (Int.) Combust., [Proc.], **21**, pp. 1579–1589.
[9] Correa, S. M., Gulati, A., and Pope, S. B., 1988, "Assessment of a Partial Equilibrium/Monte Carlo Model for Turbulent Syngas Flames," J. Cell. Biochem., **72**, pp. 159–173.
[10] Masri, A. R., and Dibble, R. W., 1988, "Spontaneous Raman Measurements in Turbulent $CO/H_2/N_2$ Flames Near Extinction," Sym. (Int.) Combust., [Proc.], **22**, pp. 607–618.
[11] Glassman, I., 1996, *Combustion*, Academic, NY.
[12] Schefer, R. W., 2003, "Hydrogen Enrichment for Improved Lean Flame Stability," Int. J. Hydrogen Energy, **28**, pp. 1131–1141.
[13] Wicksall, D., and Agrawal, A., 2001, "Effects of Fuel Composition on Flammability Limit of a Lean, Premixed Combustor," ASME Paper No. 2001-GT-0007.
[14] Sankaran, R. and Im, H. G., 2002, "Dynamic Flammability Limits of Methane/Air Premixed Flames with Mixture Composition Fluctuations," Proc. Combust. Inst., **29**, pp. 77–84.
[15] Jackson, G. S., Sai, R., Plaia, J. M., Boggs, C., Kiger, M., and Kenneth, T., 2003, "Influence of H_2 on the Response of Lean Premixed CH_4 Flames to High Strained Flows," Combust. Flame, **132**, pp. 503–511.
[16] Vagelopoulos, C. M., and Egolfopoulos, F. N., 1994, "Laminar Flame Speeds and Extinction Strains Rates of Mixtures of Carbon Monoxide with Hydrogen, Methane, and Air," Sym. (Int.) Combust., [Proc.], **25**, pp 1317–1323.
[17] Yoshimura, T., McDonell, V., and Samuelsen, S., 2005, "Evaluation of Hydrogen Addition to Natural Gas on the Stability and Emission Behavior of a Model Gas Turbine Combustor," ASME Paper No. GT2005-68785.
[18] Zukoski, E. E., 1997, "Afterburners," in *Aerothermodynamics of Gas Turbine and Rocket Propulsion*, G. Oates, ed., AIAA.
[19] Spaulding, D., 1955, *Some Fundamentals of Combustion*, Butterworth, London, Chap. 5.
[20] Longwell, J., Frost, E., and Weiss, M., 1953, "Flame Stability in Bluff-Body Recirculation Zones," Ind. Eng. Chem., **45**(8), pp. 1629–1633.
[21] Plee, S. L., and Mellor, A. M., 1979, "Characteristic Time Correlation for Lean Blowoff of Bluff Body Stabilized Flames," Combust. Flame, **35**, pp. 61–80.
[22] Radhakrishnan, K., Heywood, J., and Tabaczynski, R., 1981, "Premixed Turbulent Flame Blowoff Velocity Correlation Based on Coherent Structures in Turbulent Flows," Combust. Flame, **42**, pp. 19–33.
[23] Putnam, A. A., and Jensen, R. R., 1949, "Appplication of Dimensionless Numbers to Flash-back and Other Combustion Phenomena," Sym. (Int.) Combust., [Proc.], **3**, pp. 89–98.
[24] Hoffmann, S., Habisreuther, P., and Lenze, B., 1994, "Development and Assessment of Correlations for Predicting Stability Limits of Swirling Flames," Chem. Eng. Process., **33**, pp. 393–400.
[25] Kido, H., Nakahara, M., Nakashima, K., and Hashimoto, J., 2002, "Influence of Local Flame Displacement Velocity on Turbulent Burning Velocity," Proc. Combust. Inst., **29**, pp. 1855–1861.
[26] Lieuwen, T., Torres, H., Johnson, C., Daniel, B. R., and Zinn, B. T., 2001, "A Mechanism for Combustion Instabilities in Lean Premixed Gas Turbine Combustors," J. Eng. Gas Turbines Power, **123**(1), pp. 182–190.
[27] Natarajan, J., Nandula, S., Lieuwen, T., and Seitzman, J., 2005, "Laminar Flame Speeds of Synthetic Gas Fuel Mixtures," ASME Paper No. GT2005-68917.
[28] Kee, R. J., Warnatz, J., and Miller, Ja., 1983, "A FORTRAN Computer Code Package for the Evaluation of Gas-Phase Viscosities, Conductivities, and Diffusion Coefficients." Technical Report No. SAND89-8009, Sandia National Laboratories.
[29] Kido, H., Nakahara, M., 1998, "A Model of Turbulent Burning Velocity Taking the Preferential Diffusion Effect into Consideration," JSME Int. J., Ser. B, **41**(3), pp. 666–673.

NO_x Emissions of a Premixed Partially Vaporized Kerosene Spray Flame

Stefan Baessler[1]
e-mail: baessler@td.mw.tum.de

Klaus G. Mösl

Thomas Sattelmayer

Lehrstuhl für Thermodynamik,
Technische Unversität München,
D-85748 Garching, Germany

An important question for future aeroengine combustors is how partial vaporization influences the NO_x *emissions of spray flames. In order to address this question an experimental study of the combustion of partially vaporized kerosene/air mixtures was conducted. This assesses the influence of the degree of fuel vaporization on the* NO_x *emissions in a wide range of equivalence ratios covering the entire lean burning regime. The tests were performed at atmospheric pressure, inlet air temperatures of* 313–376 K*, a reference mean air velocity of* 1.35 m/s*, and equivalence ratios of 0.6, 0.7, and 0.9 using Jet A1 fuel. An ultrasonic atomizer was used to generate a fuel spray with a Sauter Mean Diameter of approximately* 50 μm*. The spray and the heated air were mixed in a glass tube of* 71 mm *diameter and a variable length of* 0.5–1 m*. The temperature of the mixing air and the length of the preheater tube were used for the control of the degree of vaporization. Downstream of the vaporizing section, the mixture was ignited and the flame was stabilized with a hot wire ring that was electrically heated. For local exhaust measurements a temperature controlled suction probe in combination with a conventional gas analysis system were used. The vaporized ratio of the injected fuel was determined by a Phase Doppler Anemometer (PDA). In order to optimize the accuracy of these measurements extensive validation tests with a patternator method were performed and a calibration curve was derived. The data collected in this study illustrates the effect of the vaporization rate* ψ *upstream of the flame front on the* NO_x *emissions which changes with varying equivalence ratio and degree of vaporization. In the test case with low prevaporization the equivalence ratio only has a minor influence on the* NO_x *emissions. Experiments made with air preheat and higher degrees of vaporization show two effects: With increasing preheat air temperature,* NO_x *emissions increase due to higher effective flame temperatures. However, with an increasing degree of vaporization, emissions become lower due to the dropping number and size of burning droplets, which act as hot spots. A correction for the effect of the preheat temperature was developed. It reveals the effect of the degree of prevaporization and shows that the* NO_x *emissions are almost independent of* ψ *for near-stoichiometric operation. At overall lean conditions the* NO_x *emissions drop nonlinearly with* ψ*. This leads to the conclusion that a high degree of vaporization is required in order to achieve substantial* NO_x *abatement.* [DOI: 10.1115/1.2718570]

Introduction

Combustion processes with excess air exhibit a high potential for the reduction of the NO_x emissions from spray flames in gas turbines. The concept of lean premixing has been successfully introduced for utility gas turbines operating with gaseous as well as with liquid fuels. Typically the NO_x emissions are reduced to 25–50 ppm, 15% O_2, if full prevaporization and premixing can be achieved [1].

Due to safety concerns the progress has been much slower in aeroengine combustion and a substantial NO_x reduction could not be achieved due to increasing engine pressures and temperatures. In the past the mixing intensity in the primary zone of aeroengine combustors was improved and air staging was implemented. It was shown that these methods only have limited potential to lower NO_x and new concepts have to be found. Due to safety reasons fully-premixed systems are unlikely to appear in the near future as they are prone to autoignition and flashback. Since the current emission levels of aeroengine combustors are far above the limits for the mentioned, fully premixed systems, a promising method is the partial premixing of spray. However, the potential of the method depends considerably on the relation between the degree of vaporization and the related NO_x emission. Only in the case that a low degree of vaporization already leads to substantial improvements novel premixers with sufficiently short residence times can be designed, which do not suffer from the reliability problems experienced in the past.

In an early study in 1980, Cooper [2] made experiments with gaseous kerosene premixed in air. The investigated parameters were inlet temperatures of 600 and 700 K, an inlet pressure of 0.3 MPa, a reference velocity of 35 m/s, and equivalence ratios of 0.6 and 0.72. Probably due to flame instabilities, the degree of vaporization was not investigated beyond the burning limit of the gaseous phase and the technically more interesting regime of low and moderate degrees of vaporization could not be covered. An NO_x reduction of approximately 70% was achieved for the optimum parameters. Newer investigations on a dry, low NO_x burner with staged prevaporizer/premixer were made by Malte and co-workers [3]. They investigated the lower limits of the NO_x production of a jet stirred reactor, which was operated near full vaporization. The effect of different fuel types was studied, both, experimentally and in simulation. Rizk and co-workers [4] applied a network of chemical reactor models to predict NO_x emissions of

[1]Corresponding author.

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received October 14, 2006; final manuscript received October 24, 2006. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2006: Land, Sea and Air (GT2006), May 8–11, 2006, Barcelona, Spain, Paper No. GT2006-90248.

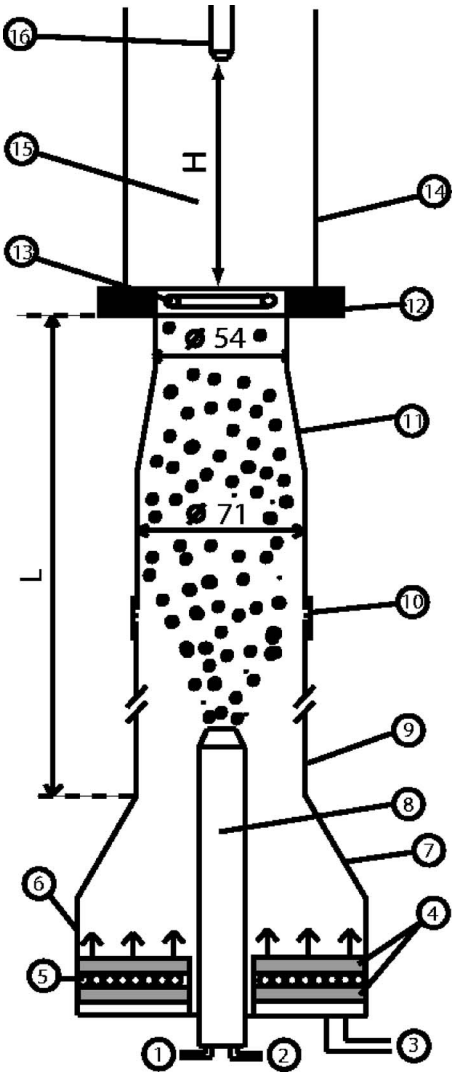


Fig. 1 Partially prevaporized spray burner (PPSB) with variable tube length *L*

aeroengines. They modelled the combustion of droplets in the primary zone with a reactor operated at stoichiometric equivalence ratio. Applying some semiexperimental correlations, they found good agreement with the experiment for several engine operation points.

This study aims at providing data to improve NO_x modeling in prevaporized, premixed spray combustion. Furthermore, this paper attempts to answer the question, how NO_x is reduced with an increased degree of vaporization covering the entire range between the nonvaporized and the fully-vaporized state.

Experimental Setup

In this section, the experimental setup is presented including the phase Doppler anemometry and patternator calibration techniques for measuring the local spray mass flow rate and spray vaporization.

Test Facility. Figure 1 shows the partially prevaporized spray burner (PPSB), in which the kerosene fuel (1) is fed to an ultrasonic nozzle (8). It uses a small mass flow of carrier air (2) (20 l_n/min) to improve the radial dispersion of the droplets. The main air (3) is heated by a set of sinter metal plates (4) with an integrated heater coil (5). During the measurements a flow rate of 300 l_n/min was used, which is substantially higher than the atomizing air flow. This main air flow enters the the vessel (6) and is accelerated by a nozzle (7) before it enters the glass tube (9), where the mixing with the spray takes place. The total length of the vaporization and mixing zone L can be adjusted by adding tube elements which are interconnected by aluminum rings (10). Downstream of the mixing zone, a second nozzle (11) is used to increase the flow velocity in the wall boundary layer in order to prevent flame flashback. The mixture is ignited by an electrically heated 1 mm wire ring (13) with 40 mm in diameter mounted in a water cooled section (12). In the wake of this ignition source a stable flame develops and spreads over the cross section further downstream. The flame is confined by a 90 mm borosilicate glass tube (14). A temperature-controlled gas probe (16) and a conventional gas analysis system are used to analyze the exhaust at the axial position H (15) above the hot wire ring. The design of the setup resulted from computational fluid dynamics and experimental optimization to prevent droplets from hitting the tube and nozzle walls. Mass flux measurements with a patternator across various tube sections indicated no significant deposition.

It is worth mentioning that measurements with a thermocouple along the tube axis showed that the energy entrained by the hot wire does not increase the temperature in the area where the measurements were taken. Thus, the increase of NO_x formation due to that effect can be neglected.

Table 1 Collision correction factors g_i

g_i	Value
Ar	–0.218
CO	–0.054
O_2	0.052
H_2	0.153
CO_2	0.476
H_2O	1.560

Gas Analysis. Exhaust gas measurements were taken on the tube axis (Fig. 1) above the hot wire ring used to ignite the mixture. The sampling probe made of stainless steel was thermally controlled with oil at 400 K to immediately stop NO forming reactions. The exhaust gas analysis system included a chemiluminescence NO_x analyzer and infrared absorption sensors for O_2, CO_2, CO, and N_2O,

$$NO + O_3 \rightarrow NO_2 + O_2 \quad 80\% \tag{1}$$

$$NO + O_3 \rightarrow NO_2^* + O_2 \quad 20\% \tag{2}$$

$$NO_2^* + M \rightarrow NO_2 + M + h\nu \tag{3}$$

$$\chi_{NO,cor} = \chi_{NO,meas} \cdot K \tag{4}$$

$$K = 1 + \sum \chi_i \cdot g_i \tag{5}$$

The measurement principle of the chemiluminescence analyzer is based on the addition of ozone (Eqs. (1) and (2)) that yields 20% of excited NO_2^*. NO is then evaluated based on the detected photons from the reaction of Eq. (3). Calibration was done with reference gases of NO, NO_2, and N_2O in N_2. Gärtner pointed out that the measured NO value has to be corrected with respect to the composition of the exhaust gas [5].

The major exhaust species CO_2, CO, H_2O, H_2, O_2, and N_2 have different individual impacts on the third body reaction taking place before the gas enters the detection chamber. For this reason, a correction has to be made, which is shown in Eq. (4). The factor K is calculated according to Eq. (5) on the basis of the exhaust gas concentrations χ_i and their relative collision factor g_i, as listed in Table 1. For complete combustion of kerosene K factors between

1.26 and 1.14 are to be employed for equivalence ratios of 1.0 and 0.5, respectively.

The equivalence ratio is calculated by Eq. (7) based on the exhaust concentrations χ_{CO_2} and χ_{CO} of the global reaction Eq. (6). CO is often neglected but it has to be accounted for rich and near stoichiometric spray flames with a low degree of vaporization. Subsequently, the NO_x emission index Eq. (8) will be used for the presentation of the emission data, which is better suited to compare the combustion efficiency than correcting NO_x concentration to a certain oxygen dilution in the exhaust [6]

$$\phi C_x H_y + a(O_2 + 3.76N_2) \rightarrow bCO_2 + cH_2O + dO_2 + eN_2 + fCO \tag{6}$$

$$\phi_{CO,CO_2} = \frac{476a(\chi_{CO} + \chi_{CO_2})}{50x(2 - \chi_{CO}) + 25y(\chi_{CO} + \chi_{CO_2})} \tag{7}$$

$$EINO_x = \frac{\chi_{NO} * M_{NO} + \chi_{NO_2} * M_{NO_2}}{\chi_{CO_2} + \chi_{CO}} \cdot \frac{x}{M_{fuel}} \tag{8}$$

Measurement of Mass Flow and Vaporization Rate. In this section the global and local degree of vaporization ψ are defined. To determine ψ experimentally a Phase Doppler Anemometer (PDA) is used. This method is known to have difficulties with the measurement of mass flux. Therefore a calibration procedure is shown in the following that improves accuracy.

The global degree of vaporization ψ is defined in Eq. (9) by the ratio of vapor mass m_v and the total fuel mass m_{tot} in the gaseous and liquid state. As it is difficult to measure the vapor mass m_v directly, Eq. (10) is introduced, which is based on the fluid mass flow rate $\dot{m}_f$ instead. To obtain $\dot{m}_f$ by means of PDA an integration of the local mass flux in relation to the PDA detection area is used in Eq. (11)

$$\psi = \left.\frac{m_y}{m_{tot}}\right|_L = \left.\frac{\dot{m}_v}{\dot{m}_v + \dot{m}_f}\right|_L \tag{9}$$

$$\psi = \left.\frac{\dot{m}_{tot} - \dot{m}_f}{\dot{m}_{tot}}\right|_L \tag{10}$$

$$\dot{m}_f|_L = {}_0^R\dot{m}''_{PDA}(r)2\pi r dr|_L \tag{11}$$

The local degree of vaporization is defined by Eqs. (12) and (13), assuming a homogenous vapor density distribution across the tube section at tube length L, with $\rho_v(r)|_L = \rho_v|_L$. The local vapor velocity $\bar{v}_v(r)|_L$ is assumed to be equal to the local air (mixture) velocity $\bar{v}_a(r)|_L = \bar{v}_{tracer}(r)|_L$.

$$\psi(r) = \left.\frac{\bar{v}_a(r) \cdot \rho_v}{\bar{v}_a(r) \cdot \rho_v + \dot{m}''_{PDA}(r)}\right|_L \tag{12}$$

$$\rho_v = \left.\frac{\dot{m}_v}{\dot{V}_a}\right|_L = \left.\frac{\dot{m}_{tot} - {}_0^R\dot{m}''_{PDA}(r)2\pi r dr}{{}_R^0\bar{v}(r)2\pi r dr}\right|_L \tag{13}$$

To date, mass flux measurements with the PDA method still require further improvement. Measurement errors are mainly caused by three effects: Two or more droplets being in the measurement volume at the same time cannot be detected correctly. The related effects are called slit and trajectory effect. Another issue is the uncertainty of the effective measurement area A_{ref} [7,8]. Several authors addressed this topic and made suggestions to improve accuracy [9–12]. New systems like the Shadow PDA [13] are able to determine whether a droplet is fully or partially within the measurement volume. The used 1D Dantec system, however, cannot acquire this information and corrections therefore have to be made. Saffman proposed to adjust the effective area for each droplet size class statistically [14]. Nevertheless a comparison of this method with patternator measurements showed significant deviations depending on the utilized PDA system [10,12].

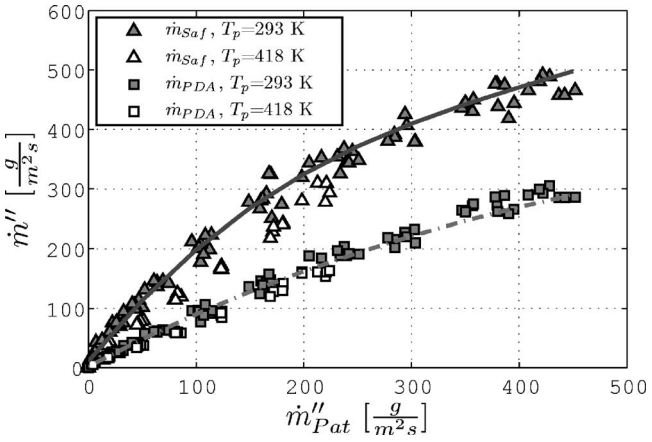


Fig. 2 Calibration curves for the PDA mass flux data using the Dantec/Saffman method and Eq. (14) (FIT)

The spray flow in the mixing tube of the experiment has simple tubular characteristics and a patternator method can be applied to calibrate the mass flux measured by the PDA. In these calibration tests, the experimental setup, which is shown in Fig. 1, was turned upside down and a small glass test-tube of 6 mm inner diameter and 15 mm height was traversed by a computer controlled servo motor over the mixing tube's cross section. The mass collected by the patternator probe was determined gravimetrically. It was shown that the systematic error in terms of residence time could be reduced to less than 1%. The standard deviation in sequential test runs was less than 5%. Figure 2 shows the calibration curve based on the mass flow calculated by the Dantec software (Version 2.12) using a method proposed by Saffman [8,14]. Drop size classes are evaluated only when they include at least 10 droplets. In vaporizing sprays, however, a few large droplets typically prevail and have to be accounted when the mass flux is evaluated. Figure 3 compares two cases of a water spray in a 293 and 418 K air stream. For the latter one the mass flux is substantially underestimated (up to 40%) for low flux and high degrees of vaporization. Therefore a different method is used to improve measurements for vaporized sprays:

A straightforward method for defining the mass flux is provided by Eq. (14). It is based on the the fluid density ρ_f, the total number of detected droplets N, the mean droplet diameter D_{30}, and a con-

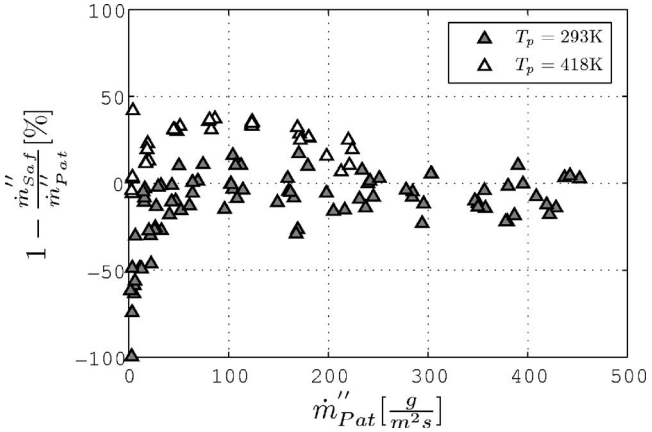


Fig. 3 Quality of the PDA mass flux data after calibration using the Dantec/Saffman method

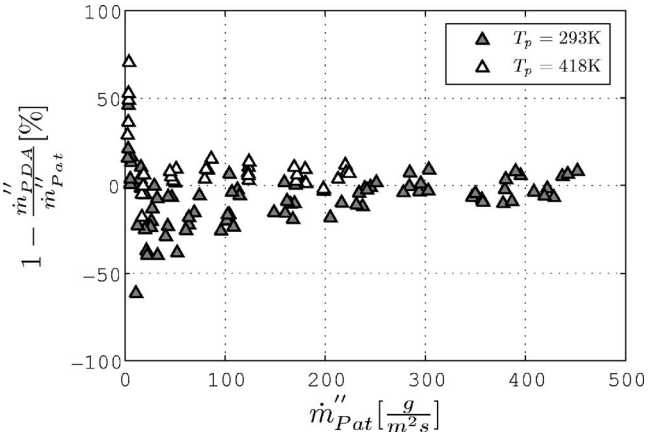


Fig. 4 Quality of the PDA mass flux data after calibration using Eq. (14)

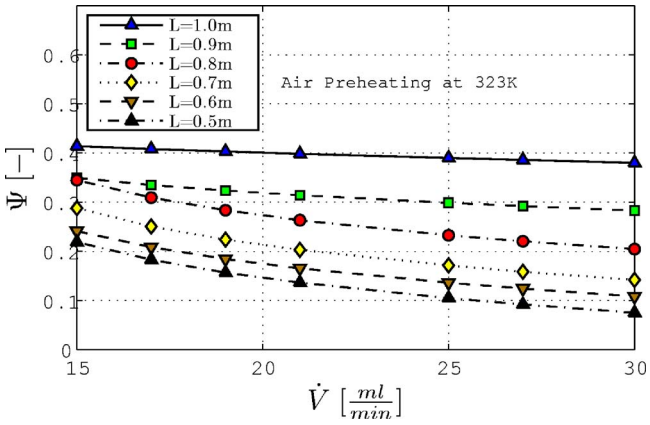


Fig. 6 Degree of vaporization for different spray flow rates $\dot{V}_k$ and tube length L

stant PDA reference area A_{ref}. With the data of several experiments, A_{ref} was calculated by solving Eq. (15). The degree of vaporization ψ is then calculated by Eq. (10).

$$\dot{m}''_{\mathrm{PDA}} = \frac{\rho_f \cdot N \cdot \frac{\pi}{6} D_{30}^3}{t_{\mathrm{PDA}} \cdot A_{\mathrm{ref}}} \tag{14}$$

$$\dot{m}_f|_L = {}_0^R \dot{m}''_{\mathrm{PDA}}(r) 2\pi r dr|_L \tag{15}$$

The calibration curve obtained by this procedure is also shown in Fig. 2. In the case of an air temperature of 418 K and fast vaporizing droplets, the deviations of the direct method become less than 20% for low mass flux and even better than 10% for high mass flux (see Fig. 4). Finally, the rather simple method with a constant reference area A_{ref} and the used calibration curve showed the best accuracy for vaporizing as well as nonvaporizing spray conditions.

Results

Fuel Mass Flux and Vaporization Rate. The local degree of vaporization was calculated with Eqs. (11) and (13). Data are plotted in Fig. 5 for a water spray flow rate of 15 ml/min at an initial 293 K in heated air of 293, 383, and 418 K. The position of the measured cross section was 0.8 m above the atomizer outlet. Even at an air temperature of 293 K vaporization takes place and 20–30% of the liquid were vaporized. The reason for this effect is the finite time, which is required in order to mix the droplets with air. A wet bulb measurement revealed a relative humidity of the process air of approximately 5% at the inlet. In particular small droplets in the spray vaporize fast and the air absorbs the water vapor. An important finding of the tests was that the degree of vaporization was found to be remarkably uniform for the inner tube area. This result justifies the assumption of quasi-one-dimensional behavior of the flow, which was adopted for the analysis of the influence of vaporization on the NO_x formation presented below.

In the experimental setup (Fig. 1), the degree of vaporization can be changed by increasing the mixing air temperature and/or increasing the mixing tube length L. The influence of both parameters on the degree of vaporization upstream of the flame was studied in the temperature range from 313 to 376 K, for a tube length between 0.5 and 1.0 m, and for volumetric kerosene spray flow rates between 15 and 30 ml/min. As an example, the results for 323 K are plotted in Fig. 6. The degree of vaporization linearly decreases with an increasing spray flow rate. A nonlinear relation can be seen with respect to the tube length L in Fig. 7.

NO_x Emissions. The emission measurements of this study covered one near stoichiometric and two lean equivalence ratios, and the entire range from nonvaporized to nearly full vaporized sprays. In the measurements the degree of vaporization ψ was varied in two ways: In each test series the tube length was varied while maintaining air preheat constant. The minimum tube length

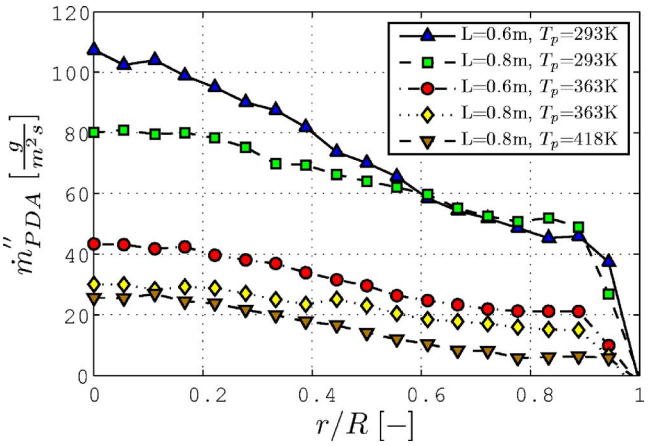


Fig. 5 Mass flux distribution by PDA at different tube lengths ($\dot{V}_w$=15 mL/min)

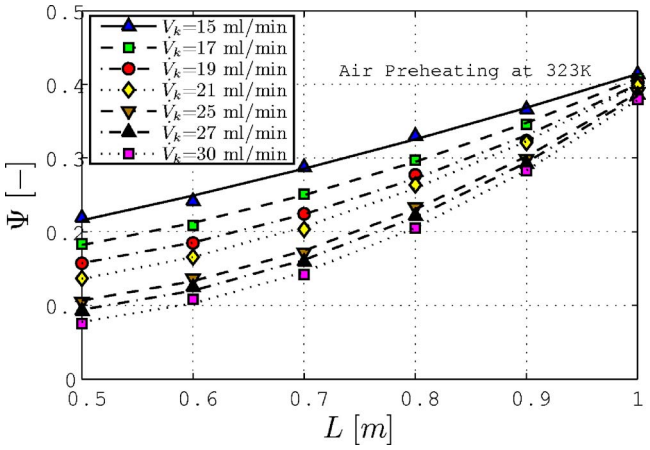


Fig. 7 Degree of vaporization for different tube lengths L and volumetric flow rate $\dot{V}_k$

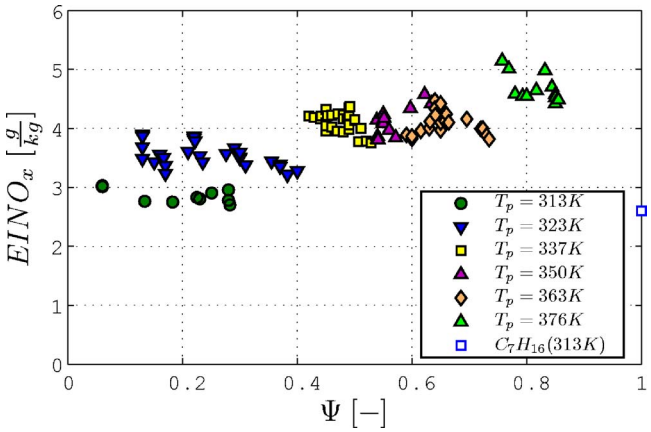


Fig. 8 Influence of the degree of vaporization ψ on the NO_x emissions (equivalence ratio 0.9)

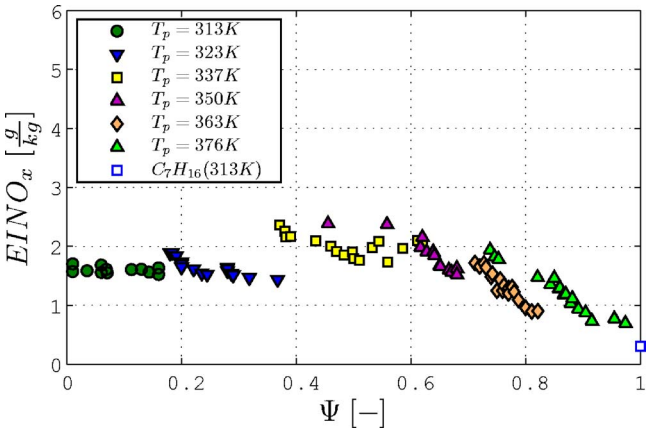


Fig. 10 Influence of the degree of vaporization ψ on the NO_x emissions (equivalence ratio 0.6)

was limited to approximately 0.5 m in order to achieve a homogeneous radial spray distribution and a uniform equivalence ratio, whereas the longest tube used was 1 m due to constraints given by the laboratory and the spray motion. To fully cover the entire range of vaporization the series of tests were repeated with six air preheat temperatures.

Figures 8–10 show the effect of the degree of vaporization on the NO_x emissions. Each data point represents a mean value of 10 s of measurement. Since a transition between spray combustion with high NO_x production and premixed gaseous combustion resulting in low NO_x emissions takes place, a substantial decrease of the NO_x emissions is expected at overall lean conditions with an increasing degree of vaporization if the influence of air preheating on the flame temperature is compensated. At near stoichiometric conditions, however, nonpremixed and premixed flames are known to emit similar quantities of NO_x. Figure 8 even reveals an NO_x increase with increasing vaporization. The reason for this effect is that higher air preheat was used in the measurements with higher degrees of vaporization leading to an increase of the maximum flame temperature and a higher NO formation. At leaner conditions (Figs. 9 and 10) the effect of air preheat is also visible in the range of low degrees of vaporization. It masks the effect of vaporization, which becomes dominant in the range of almost completely prevaporized sprays. In order to compensate the effect of a nonconstant adiabatic flame temperature, two methods can be used to calculate suitable correction factors f_i for each temperature level (Eq. (16)). The basic idea is to relate all measurements to a common air preheat temperature level T_{ref},

$$\text{EINOx}_{\text{cor}}|_{T_{\text{ref}}} = \left.\frac{\text{EINOx}}{f_i}\right|_{T_p} \tag{16}$$

A somewhat crude method for the determination of the flame temperature compensation factors is directly based on the differences of the NO_x emissions. A comparison of the data points, with an equal degree of vaporization ψ but a different air preheat, shows an incremental increase in the NO_x values. This step stems from the increase in flame temperature which is caused by the higher, succeeding preheat level. This method provides a first estimation of the correction factors to compensate the air preheat effect.

A more thorough approach is to employ reaction kinetics simulations with a detailed mechanism to study the increase of NO_x with respect to the inlet temperature [15]. For the sake of simplicity an *n*-heptane mechanism was used in Chemkin calculations, which was published by Naha and Aggarwal in 2004 [16]. They combined the *n*-heptane mechanism published by Held et al. [17] and the NO_x chemistry extension of Li and Williams [18]. This simplification appears to be acceptable because NO chemistry and emission in the lean regime do not substantially differ for various types of C_xH_y fuels [3]. Furthermore only relative values of NO_x are required for the purpose of flame temperature compensation due to air preheating.

Figure 11 shows the calculated results of the preheat influence on the NO_x formation for a laminar *n*-heptane flame with the same residence time as encountered in the experiments (in the range of

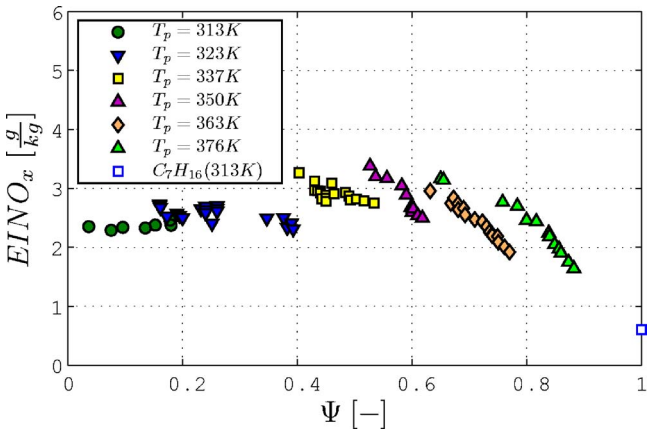


Fig. 9 Influence of the degree of vaporization ψ on the NO_x emissions (equivalence ratio 0.7)

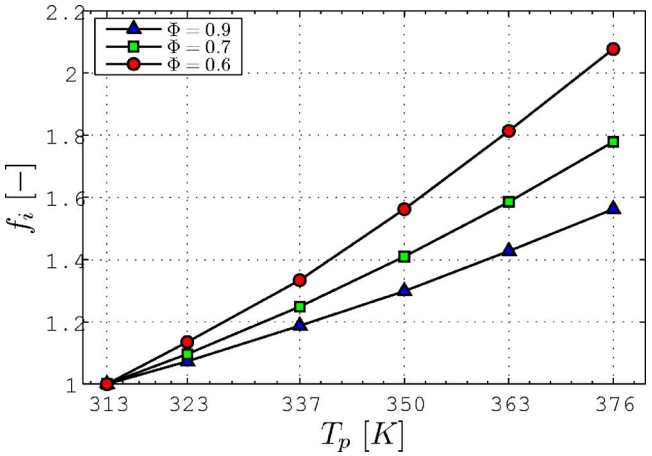


Fig. 11 Influence of the preheat temperature T_p on the NO_x emissions (laminar premixed *n*-heptane flame)

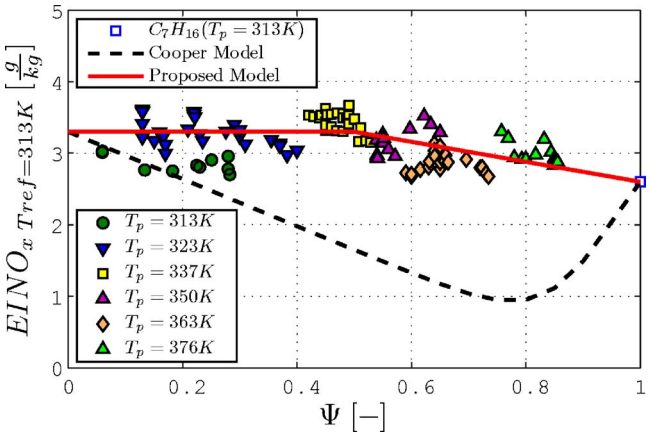


Fig. 12 Influence of the degree of vaporization ψ on the NO_x emissions corrected to T_{ref}=313 K (ϕ=0.9)

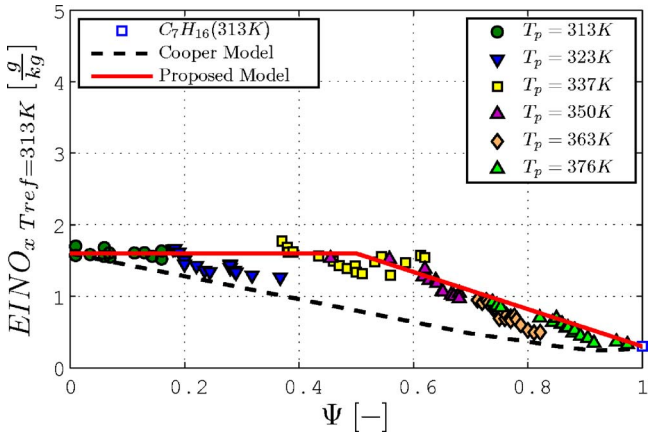


Fig. 14 Influence of the degree of vaporization ψ on the NO_x emissions corrected to T_{ref}=313 K (ϕ=0.6)

90 ms). These compensation factors were used to calculate the NO_x values for the case of no preheat according to Eq. (16). The resulting data are shown in the Figs. 12–14.

For the equivalence ratio of 0.9 the NO_x emissions remain almost constant and exhibit only a very weak drop in the regime of high degrees of vaporization. This behavior can be explained if the two asymptotic cases are compared with one another. The combustion of droplets takes place at almost stoichiometric equivalence ratios [6]. As a result of the high flame temperature in the burning region around the droplet, maximum NO_x production is encountered. If the spray is completely vaporized, the fuel will burn as a gas flame with a slightly lower effective flame temperature than before, as the gas phase has an equivalence ratio of 0.9. This leads to the conclusion that NO_x should exhibit a slight decrease over the entire range of vaporization. At the equivalence ratios of 0.7 and 0.6 the NO_x emissions generally showed a decreasing trend with ψ, as expected. In all NO_x emission plots, data scattering for constant ψ is apparent. This traces back to the fact that the highest and lowest values are associated with mixing tube lengths of 0.5 and 1.0 m, respectively. This leads to the conclusion that NO_x emission is lowered with increased tube length and, thus, mixing time.

In Figs. 12–14 the modeling approach of Cooper [2] is shown. The model is derived from the following rationale: For full vaporization NO_x only depends on the gaseous flame temperature. At partially vaporized condition a gas phase equivalence ratio can be calculated by Eq. (18) assuming a perfectly stirred gas phase. The NO_x contribution of the burned gas phase is then directly related to the flame temperature at ϕ_v. The droplets burn with a diffusion flame at an equivalence ratio of unity and near the adiabatic flame temperature. Both contributions are added and weighted with the degree of vaporization ψ in Eq. (17),

$$\mathrm{NOx}|_{\mathrm{tot}} = \psi \cdot \mathrm{NOx}|_v + (1-\psi) \cdot \mathrm{NOx}|_d \tag{17}$$

$$\phi_v = \psi \cdot \phi \tag{18}$$

In Cooper's research work model and experimental data showed reasonable agreement at high degrees of vaporization when a burnable gas phase was present. Unfortunately, measurements for the low and mid ranges of vaporization, are not shown in this study. Regarding its predicted curves, this model is not suited to match the NO_x emissions of our research work.

In order to explain the relationship between the NO_x emission and the degree of vaporization for ϕ=0.7 and 0.6, the consideration of three different zones is useful. At low degrees of vaporization ψ the emissions remain at the same level as for the non-vaporized state because the equivalence ratio of gas phase is outside of its flammability limits and the vaporized fuel is consumed within the flame zone of the droplets at ϕ=1.0. If the equivalence ratio of the gas phase is enters the regime, where gas phase combustion is possible, the fuel is consumed by a gaseous premixed flame and the NO_x contribution is low in the case of excess air conditions. This explains the low NO_x emission at high degrees of vaporization. In between these two regimes a transition region is found (Fig. 13), which already exhibits a substantial drop of the NO_x emissions although the gas phase is not yet in the flammable range.

This effect can only be explained if the assumption of the spatial uniformity of the gas phase equivalence ratio is dropped. For the time being, our working hypothesis is that small droplets are quickly vaporized in the vicinity of burning droplets or in the exhaust gas of already burnt droplets. If this is the case, the gas phase equivalence ratio ϕ_v increases locally and a combustible mixture can be formed. Furthermore, fuel vapor, which is contained in the oxidizer volume and needed for complete combustion of a single droplet, is consumed within the diffusion flame of the droplet. Fuel vapor of the unburnt area starts to react when getting into contact with the exhaust. Due to the mixing of cold unburnt mixture and exhaust, combustion takes place at lower

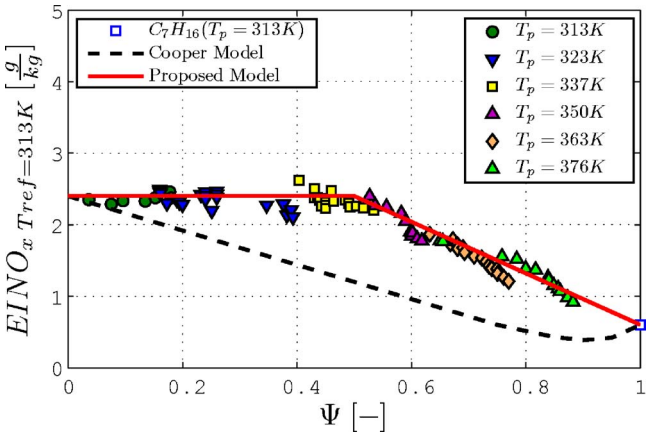


Fig. 13 Influence of the degree of vaporization ψ on the NO_x emissions corrected to T_{ref}=313 K (ϕ=0.7)

temperatures than encountered in the diffusion flame and, thus, leads to lower NO_x emissions.

For the practical point of view, an approximation of the results of the study is made in Eqs. (19) and (20). Constant NO_x emissions in the range of $0<\psi<\psi_{lim}$ and a linear drop for $\psi_{lim}<\psi<1$ with $\psi_{lim}=0.5$ are assumed. The plots of the proposed model are included in Figs. 12–14,

$$NOx|_{tot} = \gamma \cdot NOx|_v + (1-\gamma)\cdot NOx|_d \tag{19}$$

$$\gamma = \begin{cases} 0 & \psi \leq \psi_{lim} \\ \dfrac{\psi-\psi_{lim}}{1-\psi_{lim}} & \psi > \psi_{lim} \end{cases} \tag{20}$$

The onset of NO_x emission reduction at ψ_{lim} is supposed to be shifted depending on the quality of gas phase mixing. To model the mixing and vaporization effect an extended model is under development. It will be based on a single droplet NO_x model and will take various effects into account including the droplet number, droplet size distribution, and gas phase inhomogeneities using a statistical approach. After the validation of this model the pressure influence will be investigated in order to show whether the same relationship can be obtained between the relative NO_x reduction and the degree of vaporization with respect to engine pressures.

Conclusions

The influence of the degree of vaporization ψ on the NO_x emissions of a kerosene spray flame was investigated at atmospheric pressure:

- For an equivalence ratio of 0.9, NO_x emissions remained constant over a wide range of ψ and showed a slight decrease at high vaporization. NO_x emissions cannot be controlled by the increase of the degree of vaporization in the near stoichiometric regime.
- For lean conditions a reduction of NO_x emissions requires that more than half of the liquid spray mass is vaporized upstream of the flame. At low degrees of vaporization ($0<\psi<0.5$), NO_x emissions are almost constant. At high degrees of vaporization ($0.5<\psi<1$), NO_x emissions drop almost linearly.
- The lowest and highest NO_x emission for each degree of vaporization are associated with mixing tube lengths L of 1.0 and 0.5 m, respectively. Better mixing lowered the NO_x emission.
- The experimental data at 1 bar and inlet temperatures of 313–376 K showed no indication that the introduction of aeroengine fuel injectors, which provide a degree of vaporization below 50%, will efficiently reduce the NO_x emissions of the combustor. However, a comparison with Cooper's modeling [2] and experimental data at 3 bar, inlet temperatures of 600–700 K, and 35 m/s air stream velocity suggests that enhanced mixing could lower NO_x even for low and medium vaporization.

Acknowledgment

The research work was performed in the framework of the ESA Microgravity Application Program "Combustion Properties of Partially Premixed Spray Systems (CPS)" under Contract No. AO-99-094 supported by the European Space Agency (ESA) and the German Aerospace Center (DLR).

Nomenclature

A_{ref} = reference area for PDA measurement, m^2
a,b,c,d,e,f = number of moles
D_{30} = mean volumetric diameter, m
EINOx = emission index $[g(NO+NO_2)/kg(fuel)]$
f_i = correction factor
g_i = collision correction factor
H = position of probe, m
$h\nu$ = energy of emitted photon, J
K = correction factor
L = length of mixing tube, m
M = molar mass, kg/kmol
$\dot{m}$ = mass flow rate, kg/s
$\dot{m}''$ = mass flow rate per unit area, $kg/(m^2\,s)$
N = number of droplets
PDA = Phase doppler anemometer
PPSB = Partially prevaporized spray burner
R = outer tube radius, m
r = radius, m
T = temperature, K
t_{PDA} = sampling time for PDA measurement, s
$\dot{V}$ = volumetric flow rate m^3/s
v_a = velocity of air, m/s
x,y = indices in C_xH_y
χ = molar concentration
ϕ = equivalence ratio
γ = variable
ρ = density, kg/m^3
ψ = degree of vaporization
$()_a$ = air
$()_{cor}$ = corrected, after temperature compensation
$()_d$ = droplets
$()_f$ = fluid
$()_k$ = kerosene
$()_L$ = length of tube
$()_{lim}$ = limit
$()_{Pat}$ = patternator
$()_{PDA}$ = phase Doppler Anemometer
$()_p$ = air preheat
$()_{ref}$ = reference
$()_{Saf}$ = Saffman method
$()_{tot}$ = total
$()_v$ = vapor
$()_w$ = water

References

[1] Jansohn, P., Ruck, T., Steinbach, C., Knöpfel, H. -P., Troger, C., and Sattelmayer, T., 1997, "Development of the Advanced EV (AEV) Burner for the ABB GTX100 Gas Turbine," ASME Paper No. 97-AA-139.

[2] Cooper, L., 1980, "Effect of Degree of Fuel Vaporization Upon Emissions for a Premixed Partially Vaporized Combustion System," NASA Technical Paper No. 1582, Lewis Research Center, Cambridge, MA.

[3] Lee, L., Malte, P., and Benjamin, M., 2001, "Low NO_x Combustion for Liquid Fuels: Atmospheric Pressure Experiments Using a Staged Prevaporizer-Premixer," ASME Paper No. 2001-GT-0081.

[4] Rizk, N. K., et al., 2002, "Prediction of NO_x Formation Under Combined Droplet and Partially Premixed Reaction of Diffusive Flame Combustors," J. Eng. Gas Turbines Power, **124**, pp. 31–38.

[5] Gärtner, F., 1982, "Vergleich der Bildung von Stickstoffoxid in Methanol-Luft- und Kohlenwasserstoff-Luft-Flammen," Ph.D. thesis, Technische Hochschule Darmstadt.

[6] Turns, S. R., 2000, *An Introduction to Combustion*, 2nd ed., McGraw-Hill, New York.

[7] Particle Dynamics Analyzer: Supplement of the Users Manual. DANTEC Measurement Technology, Skovlunde, Denmark, 1991.

[8] BSA Flow Software v.2.1., DANTEC Measurement Technology, Skovlunde, Denmark, 2003.

[9] Qiu, H. H., and Sommerfeld, M., 1992, "A Reliable Method for Determining the Measurement Volume Size and Particle Mass Fluxes Using Phase-Doppler Anemometry," Exp. Fluids **13**, pp. 393–404.

[10] Dullenkopf, K., et al., 1998, "Comparative Mass Flux Measurements in Sprays Using a Patternator and the Phase-Doppler Technique," Part. Part. Syst. Charact., **15**, pp. 81–89.

[11] Ochs, M., 2000, "Verdunstung Monodisperser, Frei Beweglicher Brennstoff-Tropfen in Einer Turbulenten Heissluftströmung," Ph.D. thesis, ETH Zürich.

[12] Kaiser, E., 1996, "Vormischstrecke für Schadstoffarme Triebwerkskammern," Technical Report No. BMFT 2124/513-8891-LVF 9102, TU Dresden.

[13] Morikita, H., et al., 1998, "Application of Shadow Doppler Velocimetry to

Paint Spray: Potential and Limitations in Sizing Optically Inhomogeneous Droplets," Meas. Sci. Technol., **9**, pp. 221–231.
[14] Saffman, M., 1987, "Automatic Calibration of LDA Measurement Volume Size," Appl. Opt., **26**(13), pp. 2592–2597.
[15] Sattelmayer, T., et al., 1998, "NO_x-Abatement Potential of Lean-Premixed GT Combustors," ASME J. Eng. Gas Turbines Power, **120**, pp. 48–59.
[16] Naha, S., and Aggarwal, S. K., 2004, "Fuel Effects on NO_x Emissions in Partially Premixed Flames," Combust. Flame, **139**, pp. 90–105.
[17] Held, T. J., Marchese, A. J., and Dryer, F. L., 1997, "A Semi-Empirical Reaction Mechanism for *n*-Heptane Oxidatation and Pyrolysis," Combust. Sci. Technol., **123**, pp. 107–146.
[18] Li, S. C., and Williams, F. A., 1999, "NO_x Formation in Two-Stage Methane-Air Flames," Combust. Flame, **118**, pp. 399–414.

A. El-Shafei
T. A. F. Hassan
A. K. Soliman
Y. Zeyada

RITEC,
New Maadi,
Cairo 11435, Egypt

N. Rieger
STI Technologies, Inc.,
Rochester, NY 14623

Neural Network and Fuzzy Logic Diagnostics of 1x Faults in Rotating Machinery

In this paper, the application of neural networks and fuzzy logic to the diagnosis of faults in rotating machinery is investigated. The learning-vector-quantization (LVQ) neural network is applied in series and in parallel to a fuzzy inference engine, to diagnose 1x faults. The faults investigated are unbalance, misalignment, and structural looseness. The method is applied to a test rig (Hassan et al., 2003, ASME Paper No. GT 2003-38450), and the effectiveness of the integrated Neural Network and Fuzzy Logic method is illustrated. [DOI: 10.1115/1.2227417]

Introduction

Maintenance management of industrial facilities is a major task for all manufacturing companies. Investments of billions of dollars in capital equipment have to be maintained to preserve company assets and to ensure continuous production [1,2]. This fact has been recognized for a long period of time, but lately it became apparent that it is necessary for a maintenance management program to also reduce inflated maintenance costs and to improve operator safety [3–5].

Condition-based maintenance (CBM) has evolved to be the leading methodology to effectively plan for the required maintenance activities. This is the premise of the technologies of predictive maintenance [1], where the maintenance or repair action is based on the condition of the machine, i.e., only needed maintenance work is done, thus reducing maintenance costs and improving productivity [6]. However, for this technique to be effective, machinery condition should be measured and competently analyzed to support the appropriate maintenance decision [3,7,8].

Currently, in order to apply CBM, the condition of plant assets is determined by human experts, with the support of technologies such as vibration spectral measurements. The expert opinion is then somehow incorporated in the maintenance management system. Several attempts at automating the spectral analysis process by rule-based expert systems and inference engines in the late 1980s and early 1990s provided some products that are available in the market for providing recommendations mainly based on spectral analysis. These products are available from most condition monitoring equipment manufacturers, but were not quite successful in the market, because most of the time the user can provide a better recommendation than the expert system [1].

Possibilities exist nowadays for replacing the rigid rule base of expert systems with fuzzy logic, which would have the advantage of approaching human thinking. Moreover, the current availability of great computational capabilities should also be exploited. This is the objective of this paper. Actually, a condition monitoring professional may inspect the frequency spectrum of the vibration signal and decide whether the machine is in normal operation or a specific fault is in progress. In doing so, both numerical analysis represented by the equipment used to analyze the signal, and logic analysis represented by the human experience and reasoning, are used throughout the diagnostic process. It is thus clear that the process relies on the availability of the numerical capabilities as well as the professional analysts to perform the diagnosis task. This is a limitation for many industrial facilities. Issues like availability, cost, and reliability of the process should be investigated before employing a predictive maintenance program. The effort in this paper is directed towards the automation of the diagnostic procedure to incorporate both numerical and logic analyses, which minimize the need of human expertise that can be transmitted to the logic module of the diagnostic system [9].

This paper focuses on the advantage of the relatively modern techniques such as neural networks and fuzzy logic systems to improve the rotor-bearing fault diagnostic procedure. Using both techniques in a hybrid system should have the advantage of the fast and efficient numerical capabilities of neural networks as well as the flexibility and the logic inference of fuzzy systems [10,11].

The neural network provides suggestions about the type of fault suspected, while a fuzzy logic module is designed to focus on the amplitude spectrum characteristics of that particular fault, both as a stand-alone diagnostic tool and as an in-series diagnostic tool with the neural network.

Introducing logic in the diagnostic process gives the procedure the power of fast and complex computations represented by the neural network performance, in addition to the flexibility and reasoning of the fuzzy systems. This configuration suits the diagnostic procedure of a vibration specialist who uses equipment for data processing and reasoning resulting from her/his experience.

Background

A neural network (NN) is a biologically inspired computational model that consists of processing elements (neurons) and connections between them, as well as of training and recall algorithms. The structure of a neuron is defined by inputs, having weights bound to them; and an input function, which calculates the aggregated net input signal to a neuron coming from all its inputs; and an activation (signal) function, which calculates the activation level of a neuron as a function of its aggregated input signal and (possibly) of its previous state. An output signal equal to the activation value is emitted through the output of the neuron. NNs are owing to the main role of the connections. The weights bound to them are a result of the training process and represent the "long-term-memory" of the model [12]. The main characteristics of a NN are:

- Learning: a NN can start with no knowledge and can be trained using a given set of data examples, that is, input-output pairs (as in supervised training) or only input data (as in unsupervised training). Through learning the con-

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received October 1, 2005; final manuscript received February 1, 2006. IGTI Review Chair: R. S. Abhari. Paper presented at the ASME Turbo Expo 2006: Land, Sea, and Air, Barcelona, Spain, May 8–11 2006, Paper No. GT2005-68885.

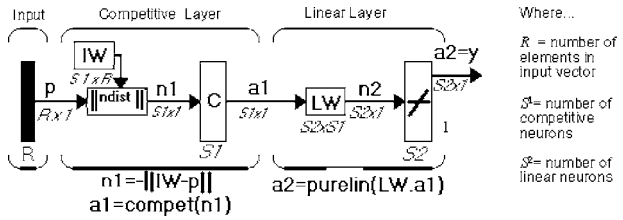


Fig. 1 LVQ neural network

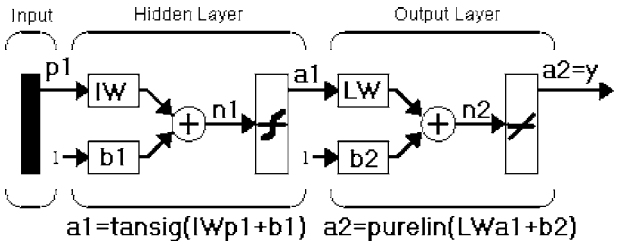


Fig. 2 Feed forward network

nection weights change in such a way that the network learns to produce desired outputs for known inputs.

- Generalization: if a new vector that differs from the known examples is supplied to the network, it produces the best output according to the examples used.
- Massive potential parallelism: during the processing of data many neurons "fire" simultaneously.
- Robustness: if some neurons go wrong, the whole system may still perform well.

One way to represent inexact data and knowledge, closer to humanlike thinking, is to use fuzzy rules instead of exact rules when representing knowledge. Fuzzy systems are rule-based expert systems based on fuzzy rules and fuzzy inference. Fuzzy rules represent in a straightforward way the common sense knowledge and skills, or knowledge that is subjective, ambiguous, vague, or contradictory [11,13]. Common sense knowledge may have been acquired from long term experience, from the experience of many people, over many years. A fuzzy system is defined by three main components:

1. fuzzy input and output variables,
2. a set of fuzzy rules, and
3. fuzzy inference mechanism.

Neural Networks Architecture

The availability and customization of the NN to fit the problem at hand was investigated in our previous work [14]. Five networks were investigated (perceptrons, linear, feed forward, self organizing, and learning-vector-quantization (LVQ)), and two networks of those previously investigated were chosen as promising networks (feed forward and LVQ) [14]. In this current work, the procedure and developed standard techniques for the diagnosis process using neural networks are extended.

The LVQ neural network provided the best performance, as demonstrated from our previous experiments [14] (see Fig. 1). Therefore we selected the LVQ network to be used in our elaborate tests as a diagnostic tool. However, the LVQ network has a shortcoming: it cannot distinguish multiple faults. As it uses a competitive layer in its input layer so it could not classify any combined faults (more than one fault in the same amplitude spectrum), there is always only one winner [15,16]. Each type of fault is represented by one neuron in the competitive layer; the characteristic of the competitive layer is that all the neurons' output is equal to zero except only one neuron output is equal to one, which is the winner neuron and represents a single fault.

Due to the incapability of the LVQ network to detect more than one fault, we used also the feed forward network due to its high flexibility to detect and classify more than one fault in the same amplitude spectrum [17]. However, the performance of the feed-forward network was modest in our earlier experiments [14]. Therefore to enhance the performance of the feed forward network (Fig. 2) we implemented some modifications to its previously used architecture, as follows:

- Two hidden layers were used instead of one with a large number of hidden neurons [14,17].
- The output layer function was changed from pureline function to logsig function and the output layer neurons were augmented from one neuron to four neurons [14,15].
- The training algorithm was changed from the Levenberg-Marquedt (LM) training algorithm (trainLM) to the Resilient Propagation (RP) training algorithm (trainRP), which is much faster, especially with a large number of neurons [14,18].

New Sets of Experiments

In order to satisfy the requirements of the training and testing of the neural network fuzzy logic parallel and series configurations a new set of experiments was carried out. Twelve sets of experiments were implemented; this was carried out by planting each fault separately first into the test rig. Each set was for a different fault or fault level; nine sets of vibration signatures of the machine train were acquired for each fault class separately.

Measurements were performed on the five-disc test rig [14]. A Bruel & Kjaer (B&K) 2526 data collector was used. Data acquisition is handled by Sentinel software in which measurements are analyzed and, if satisfactory, stored in a Matlab-friendly format ready for the application of the neural network or fuzzy logic application. The measurement process is illustrated in Fig. 3. The following are selective data settings [19]:

- Test rig speed: 1800 rpm.
- No. of samples: 400 sample.
- Amplitude unit: mm/sec rms (velocity).
- $$\text{Frequency interval } \Delta F = \frac{\text{Windowing factor} * \text{Frequency span}}{\text{Number of lines}} = \frac{1.5 * 200}{400} = 0.75 \text{ Hz}$$
- Resolution $= 2 * \Delta F = 1 * 0.75 = 1.5$ Hz

The following setup for the B&K 2526 data collector was used [19]:

- For overall measurements:

  High pass filter=10 Hz
  Averaging time=5 s
  Detector rms
  Integration one (acceleration to velocity)
  Upper limit=1 KHz
  Transducer (accelerometer) sensitivity 1 Pc/m/s^2
  Amplitude scaling=linear

- Auto spectrum measurements:

  Frequency span=200 Hz
  High pass filter=3 Hz
  Averages=4 times

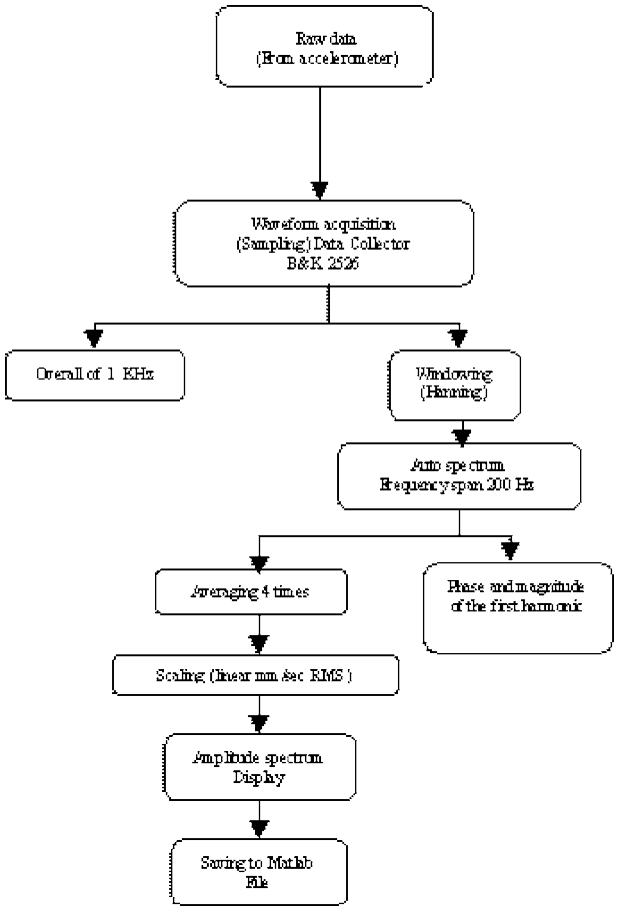


Fig. 3 Flow chart of measurement procedure

Averaging mode=spectrum
Windowing mode=Hanning window
Detector RMS
Integration one (acceleration to velocity)

Fault Planting. Neural networks and fuzzy logic should be trained and tested by data obtained by planting faults that need to be later identified by the LVQ, feed forward networks [16], and the Sugeno fuzzy inference system [20,21]. Twelve cases are considered here, no-fault, mechanical imbalance (two levels), structure looseness (two levels), bearing looseness (two levels), angular misalignment (two levels), parallel misalignment (two levels), and combined misalignment (one level). See Fig. 4 for fault implementation on test rig.

Acquired Data Sets. Nine data sets were collected for each of the previous discussed 12 classes. Each data set consists of 11 data files; each data file represents 400 samples of the amplitudes of a 200 Hz amplitude spectrum measured at a certain point on the machine train. Each data file is a vector of 1 column and 400 rows; these data files were transferred to MatLab files so they can be handled by the neural networks and the fuzzy inference system. Each set of 11 data files was measured at a predetermined measurement point location and direction on the machine train as follows:

1. Motor non drive end bearing horizontal direction (MNDE-H)
2. Motor non drive end bearing vertical direction (MNDE-V)
3. Motor drive end bearing horizontal direction (MDE-H)
4. Motor drive end bearing vertical direction (MDE-V)
5. Motor drive end bearing axial direction (MDE-A)

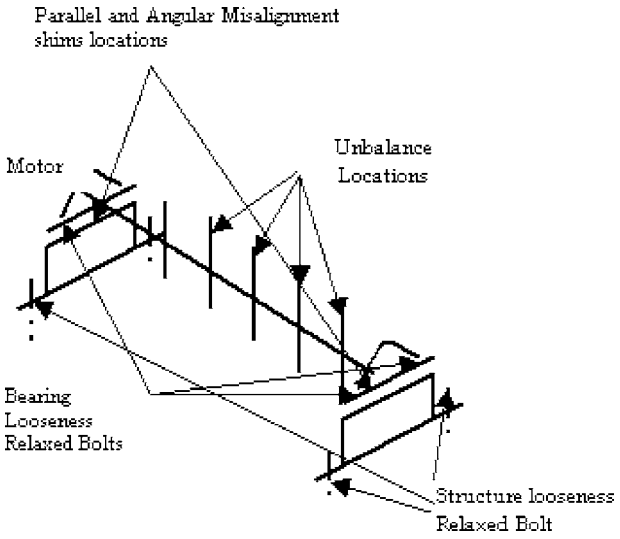


Fig. 4 Fault implementation on test rig

6. Drive end bearing horizontal direction (DE-H)
7. Drive end bearing vertical direction (DE-V)
8. Drive end bearing axial direction (DE-A)
9. Non drive end bearing horizontal direction (NDE-H)
10. Non drive end bearing vertical direction (NDE-V)
11. Non drive end bearing axial direction (NDE-A)

Files are split into two groups; the first group is used for training the networks, while a second group is used for validation of the trained networks. This technique is common in designing neural networks where the network performance is tested by data that have not been used for training to guarantee generalization.

Neural Networks Training and Testing Results

Each of the LVQ and the feed forward networks were trained and tested for validation by different data files; the files used for validation were not used for the training of the networks. The following results were obtained for each network:

LVQ Network

Total training performance: 100% correctly trained
No fault training results: 100% correctly trained
Unbalance training results: 100% correctly trained
Looseness training results: 100% correctly trained
Misalignment training results: 100% correctly trained

Total validation performance: 100% correctly identified
No fault validation results: 100% correctly identified
Unbalance validation results: 100% correctly identified
Looseness validation results: 100% correctly identified
Misalignment validation results: 100% correctly identified

It should be noted that training performance goal was met after 150 training epochs (see Fig. 5). The maximum number of epochs for training that was permitted to achieve the goal was set to 500 epochs, so the network reached its goal within the permitted number of epochs. This maximum number of epochs can be adjusted. The network succeeded in reaching its performance goal within a reasonable number of epochs.

Feed Forward Network

Total training performance: 100% correctly trained
No Fault training results: 100% correctly trained
Unbalance training results: 100% correctly trained
Looseness training results: 100% correctly trained
Misalignment training results: 100% correctly trained

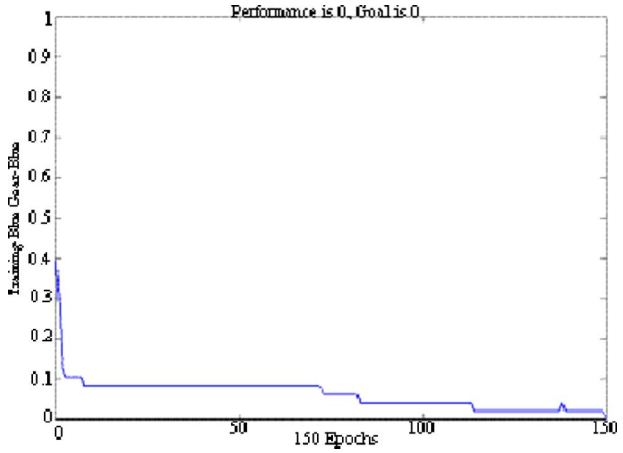


Fig. 5 Training of LVQ network

Total validation performance: 94.7% correctly identified
No Fault validation results: 100% correctly identified
Unbalance validation results: 100% correctly identified
Looseness validation results: 75% correctly identified
Misalignment validation results: 100% correctly identified

Note that the training performance goal was not met, training stopped after 27 training epochs due to reaching the minimum gradient (see Fig. 6).

The above results clearly show the success of the neural networks both in training (to become capable of diagnosing) and of really doing the diagnosis (validation). In fact, because the training and validation data sets are independent, and because our implementation of all faults at two different levels (see previous section), it is quite remarkable the success of the neural networks, particularly the LVQ Network, in identifying the faults from an array of possible faults and fault levels.

Fuzzy Logic Applied to Diagnostic Systems

Fuzzy sets were introduced by Zadeh in 1965 to represent or manipulate data and information possessing nonstatistical uncertainties [22]. It was specifically designed to mathematically represent uncertainty and vagueness and to provide formalized tools for dealing with the imprecision intrinsic to many problems [23]. Fuzzy logic provides an inference morphology that enables approximate human reasoning capabilities to be applied to knowledge-based systems. The theory of fuzzy logic provides a

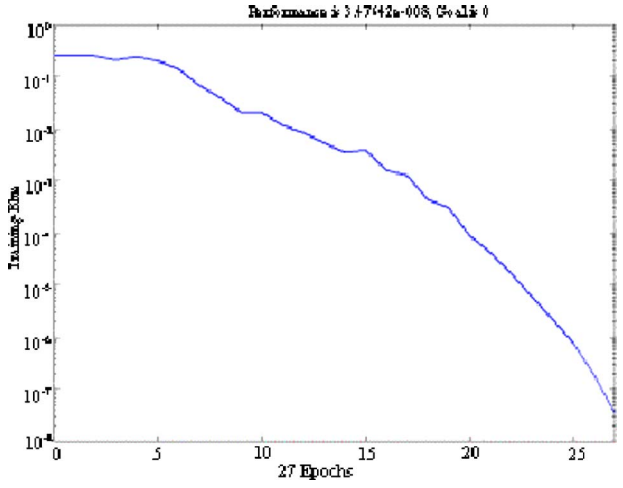


Fig. 6 Training of feed forward network

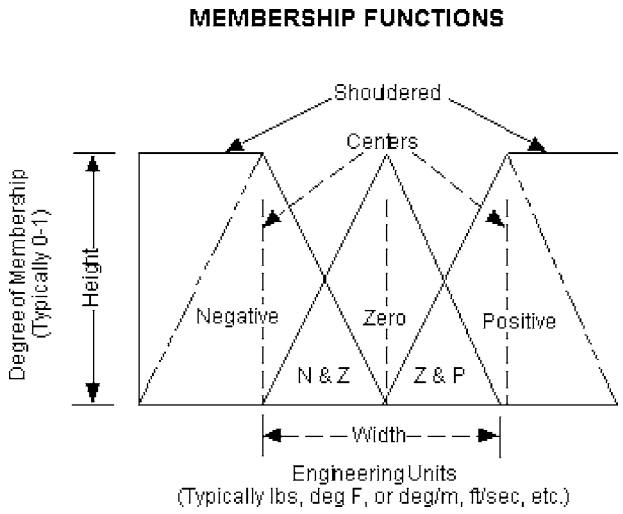


Fig. 7 The features of a membership function [20]

mathematical strength to capture the uncertainties associated with human cognitive processes, such as thinking and reasoning [24].

Definition of a Fuzzy Set. Fuzzy set A in X is characterized by its membership function

$$\mu A: X \rightarrow [0,1]$$

$\mu A(x)$ is interpreted as the degree of membership of element x in fuzzy set A for each $x \in X$.

The membership function is a graphical representation of the magnitude of participation of each input. It associates a weighting with each of the inputs that are processed, defines functional overlap between inputs, and ultimately determines an output response. The rules use the input membership values as weighting factors to determine their influence on the fuzzy output sets of the final output conclusion [22]. Once the functions are inferred, scaled, and combined, they are defuzzified into a crisp output that drives the system. There are different membership functions associated with each input and output response. Some features to note are [20]:

1. Shape: triangular is common, but bell, trapezoidal, haversine, and exponential have been used. More complex functions are possible but require greater computing overhead to implement.
2. Height or magnitude (usually normalized to 1).
3. Width (of the base of function).
4. Shouldering (locks height at maximum if an outer function, shouldered functions evaluate as 1.0 past their center).
5. Center points (center of the member function shape).
6. Overlap (N&Z, Z&P, typically about 50% of width but can be less).

Figure 7 illustrates the features of the triangular membership function, which is used in this example because of its mathematical simplicity.

Logical Operators. If we keep the fuzzy values at their extremes of 1 (completely true) and 0 (completely false), standard logical operations will hold. Now remembering that in fuzzy logic the truth of any statement is a matter of degree, the input values can be real numbers between 0 and 1. That is, resolve the statement A AND B, where A and B are limited to the range $(0,1)$, by using the function $\min(A,B)$. Using the same reasoning, we can replace the OR operation with the max function, so that A OR B becomes equivalent to $\max(A,B)$. Finally, the operation NOT A becomes equivalent to the operation $1-\mu A$.

If-Then Rules. Fuzzy sets and fuzzy operators are the subjects and verbs of fuzzy logic [24]. If-then rule statements are used to formulate the conditional statements that comprise fuzzy logic. A single fuzzy if-then rule assumes the form if x is A, then y is B, where A and B are linguistic values defined by fuzzy sets on the ranges (universes of discourse) X and Y, respectively. An example of such a rule might be If Temperature is "High," then Fan Speed is "Fast." Note that "High" is represented as a number between 0 and 1, and so the antecedent is an interpretation that returns a single number between 0 and 1. The consequent set will later be defuzzified, assigning one value to the output. System decision mainly depends on values of the system inputs and the triggered rule base [20].

Fuzzy Expert Systems. The rules in a fuzzy expert system are usually of a form similar to the following:

if x is low and y is high then z=medium

A typical fuzzy expert system has more than one rule. The entire group of rules is collectively known as a rule-base or knowledge-base [20].

The Inference Process. With the definition of the rules and membership functions in hand, we now need to know how to apply this knowledge to specific values of the input variables to compute the values of the output variables. This process is referred to as inferencing. In a fuzzy expert system, the inference process is a combination of four subprocesses: fuzzification, inference, composition, and defuzzification. The defuzzification subprocess is optional.

Fuzzification. In the fuzzification subprocess, the membership functions defined on the input variables are applied to their actual values, to determine the degree of truth for each rule premise.

Inference. In the inference subprocess, the truth value for the premise of each rule is computed and applied to the conclusion part of each rule. This results in one fuzzy subset to be assigned to each output variable for each rule.

Two typical inference methods or inference rules are MIN and PRODUCT.

Composition. In the composition subprocess, all of the fuzzy subsets assigned to each output variable are combined together to form a single fuzzy subset for each output variable.

Two composition rules are mentioned here: MAX composition and SUM composition. SUM composition is only used when it will be followed by a defuzzification method, such as the CENTROID method, that does not have a problem with this odd case [20].

Defuzzification. Sometimes it is useful to just examine the fuzzy subsets that are the result of the composition process, but more often, this _fuzzy value_ needs to be converted to a single number—a _crisp value_. This is what the defuzzification subprocess does.

Two of the more common techniques are the CENTROID and MAXIMUM methods. In the CENTROID method, the crisp value of the output variable is computed by finding the variable value of the center of gravity of the membership function for the fuzzy value. In the MAXIMUM method, one of the variable values at which the fuzzy subset has its maximum truth value is chosen as the crisp value for the output variable [20].

Diagnostics Using Fuzzy Logic

The diagnostic process of rotating machinery through vibration analysis has several tools. The most convenient technique is the spectral analysis [3].

The amplitude spectrum or simply spectrum is a plot that represents the amplitude of each frequency component in the periodic wave. The frequency components in the spectra are usually related to the frequency of rotation when dealing with rotating machinery. The running speed component is named as 1x and the second harmonic is 2x, and so on [7].

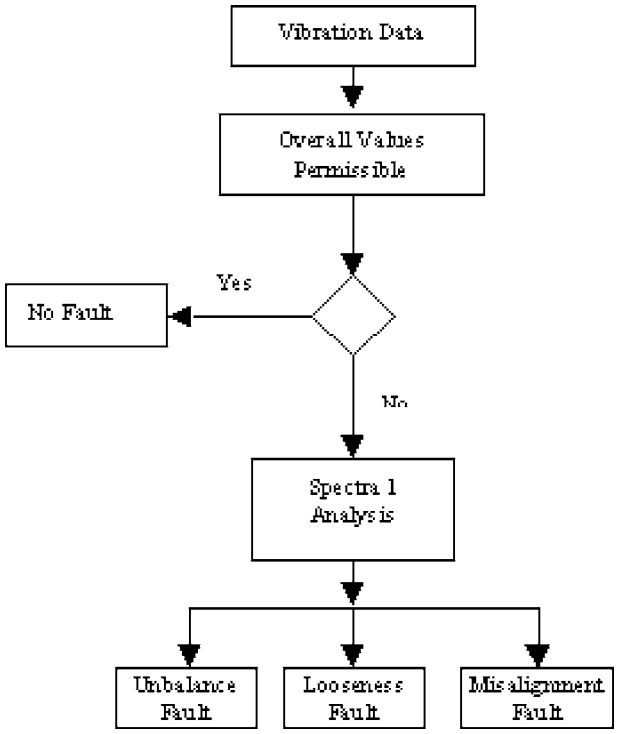


Fig. 8 Diagnosis procedure

This research is concerned with the diagnosis of three faults, namely unbalance, looseness, and misalignment. The unbalance is defined as the mass centerline is shifted from the geometrical center of the rotor. This may occur during the manufacturing as in casting process or due to wear of components. The unbalance is usually diagnosed form the spectrum by the presence of a high 1x component in the radial directions. In special cases such as in overhung rotors, the unbalance may cause high 1x vibration in the axial direction [3].

The misalignment fault is defined as the centerlines of the rotors of the driver and driven parts are not aligned together. This is a very common installation problem for rotating machinery. The misalignment is diagnosed from the spectrum by the presence of high 1x, 2x, and sometimes 3x components [3].

Looseness problems occur when the bolts that hold the bearing housing to the skid or the bolts that hold the bearing upper half to its lower half have been loosen, which is called structural looseness. In addition, when the clearances between the bearing outer race and the housing or between the bearing inner race and the shaft increased, the condition of bearing looseness occurs. The looseness can be identified from the spectrum by the presence of the 1x component with several harmonics $(2x, 3x, 4x, \ldots)$. Sometimes the $\frac{1}{2}$x component appears with its harmonics [3].

The process of diagnosing the rotating machinery faults through vibration has several steps. In the beginning, the machine should be classified according to one of the international standards according to its characteristics (load, speed, foundation,…, etc.). The aim of this classification is to judge the machine condition, whether it is in an acceptable or an unacceptable condition. If the machine is in unacceptable condition, the analysis process should then be conducted in order to determine the machine fault(s). The following chart (Fig. 8) represents the algorithm of the diagnostic procedure used in this project.

The analysis process is usually performed by human intervention to judge the measurements. In order to perform this process

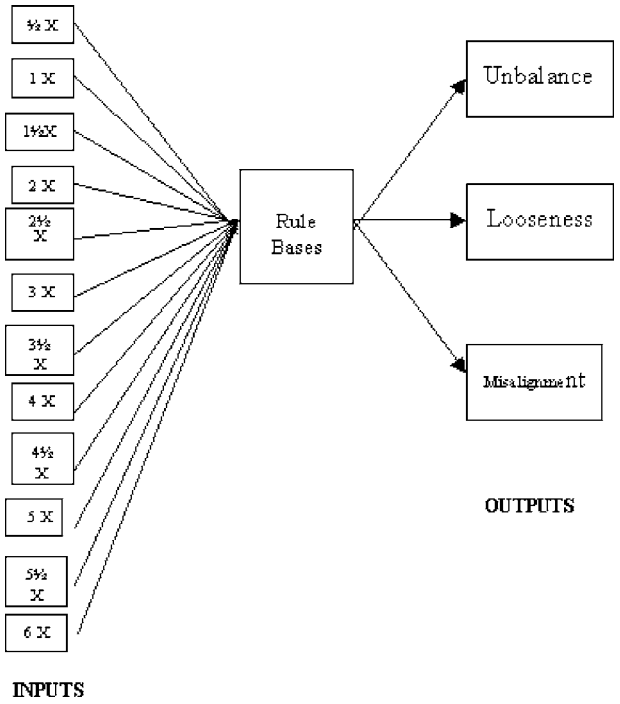


Fig. 9 Fuzzy logic system

automatically, a flexible rule set should be used, via fuzzy inference system; this is the aim of using fuzzy logic to simulate the human reasoning, because the fuzzy logic does not have a black and white decision, it has a gray scale. This is done by using the fuzzification and defuzzification process and using various types of membership functions. This flexibility is clearly found in the fuzzy logic systems. In addition, fuzzy logic has the merit of being built upon expert experience, which is the most important issue in dealing with vibration diagnosis process [25].

Table 1 Membership functions

Membership function no.	Membership function name	Left leg	Peak	Right leg
1	Very small	–0.25	0	0.25
2	Small	0	0.25	0.5
3	Medium	0.25	0.5	0.75
4	High	0.5	0.75	1
5	Very high	0.75	1	1.25

Table 3 Output results

Membership function no.	Membership function name	Value
1	No	0
2	Not sure	0.5
3.	Yes	1

Figure 9 illustrates the fuzzy logic system built for the identification of the unbalance, looseness, and misalignment faults. The system simply consists of 12 inputs representing the amplitude of the $\frac{1}{2}x, 1x, 1\frac{1}{2}x, \ldots, 6x$ components. These components are sufficient to diagnose the required three faults as described before. Thus this represents the low resolution configuration used for fuzzy diagnostics versus the high resolution configuration used for the neural network diagnosis [26], which relied on 400 lines of data in each spectrum.

Each input has the following properties:

Range: from 0 to 1
No. of membership functions: 5
Type of membership functions: triangular (trimf)

The membership functions are listed in Table 1. The rule bases are then applied in order to diagnose the fault. The rule base is simply an (If) statement that assigns a value of the membership function of each input and relates them by a logical operator, either (AND) or (OR) operator. Table 2 summarizes the rule bases built for the diagnosis process for the parallel configuration, which includes our expert knowledge. The output of the rule base is then produced indicating which of the three faults is found (Table 3).

Each output has the following properties:

Range: from 0 to 1
No. of membership functions: 3
Type of membership functions: constant

Note that we have used the Sugeno type inference system due to its suitability to this application [21].

Fuzzy Logic Results for Stand-Alone or Parallel Configuration

We have used the above system for fuzzy logic both as a diagnostic stand-alone system (parallel configuration), as well as in a

Table 2 Rule basis of the parallel configuration (VS: very small, S: small, M: medium, H: high, VH: very high, NO: none, N: no, Y: yes)

Inputs													Logical	Outputs		
	$\frac{1}{2}x$	1x	$1\frac{1}{2}x$	2x	$2\frac{1}{2}x$	3x	$3\frac{1}{2}x$	4x	$4\frac{1}{2}x$	5x	$5\frac{1}{2}x$	6x	Operator	UN	LO	MA
	VS	VH	VS	VS	VS	VS	VS	VS	VS	VS	VS	VS	AND	Y	N	N
	S	VH	S	S	S	S	S	S	S	S	S	S	AND	Y	N	N
	VS	H	VS	VS	VS	VS	VS	VS	VS	VS	VS	VS	AND	Y	N	N
	S	H	S	S	S	S	S	S	S	S	S	S	AND	Y	N	N
	M	M	M	M	M	M	M	M	M	M	M	M	AND	N	Y	N
	H	H	H	H	H	H	H	H	H	H	H	H	AND	N	Y	N
IF	VH	VH	VH	VH	VH	VH	VH	VH	VH	VH	VH	VH	AND	N	Y	N
	NO	NO	NO	NO	NO	NO	NO	VH	NO	VH	NO	VH	OR	N	Y	N
	NO	NO	NO	NO	NO	NO	NO	H	NO	H	NO	H	OR	N	Y	N
	VS	VH	VS	VH	VS	VS	VS	VS	VS	VS	VS	VS	AND	N	N	Y
	VS	H	VS	H	VS	VS	VS	VS	VS	VS	VS	VS	AND	N	N	Y
	VS	M	VS	VH	VS	VS	VS	VS	VS	VS	VS	VS	AND	N	N	Y
	VS	S	VS	VH	VS	VS	VS	VS	VS	VS	VS	VS	AND	N	N	Y
	VS	M	VS	H	VS	VS	VS	VS	VS	VS	VS	VS	AND	N	N	Y
	VS	S	VS	H	VS	VS	VS	VS	VS	VS	VS	VS	AND	N	N	Y

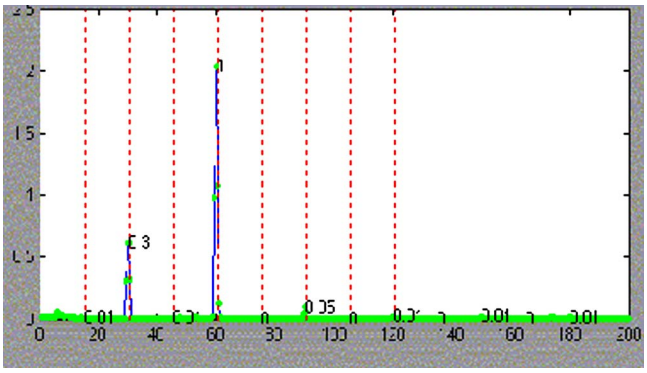


Fig. 10 Misalignment sample amplitude spectrum

series system with neural network as a first step and fuzzy logic as a second step, to provide a confidence level in the neural network diagnosis. The data used for applying the fuzzy logic system are those collected for the neural network diagnosis, as described earlier in this paper. The data collected represented the cases of no fault, unbalance fault, looseness fault, and misalignment fault, as described earlier (Fig. 10).

Applying the fuzzy logic diagnostic system rule basis of the parallel configuration to the collected data, the diagnostic results are summarized as follows.

Total cases studied: 171
No. of cases correctly identified: 168
Percentage: 98%

No fault cases studied: 99
No. of cases correctly identified: 99
Percentage: 100%
Unbalance cases studied: 18
No. of cases correctly identified: 18
Percentage: 100%
Looseness cases studied: 36
No. of cases correctly identified: 33
Percentage: 92%
Misalignment cases studied: 18
No. of cases correctly identified: 18
Percentage: 100%

The results indicate excellent performance of the built fuzzy logic system. Only 1 data file out of 171 data files introduced to the system was identified incorrectly and 2 files were not specified.

The system was able to identify the no fault, unbalance, and misalignment cases without errors, and three errors in diagnosing the looseness.

Series Configuration

The series configuration depends on developing fuzzy logic inference engine to provide a confidence level for neural networks identified cases. This technique is referred to as series configuration or series fuzzy inference (SFI). This is an application of neuro-fuzzy diagnostics [26]. Experimental data were used to test the fuzzy system and to evaluate the performance of the fuzzy inference rules.

A normalized amplitude vector is extracted from each spectrum corresponding to whole and half harmonics up to 6x. These values

Table 4 Unbalance rule basis for SFI (VS: very small, S: small, M: medium, H: high, VH: very high, L: low)

Logical operator		$\frac{1}{2}$x	1x	$1\frac{1}{2}$x	2x	$2\frac{1}{2}$x	3x	$3\frac{1}{2}$x	4x	$4\frac{1}{2}$x	5x	$5\frac{1}{2}$x	6x		Confidence Level
AND	If	VS	VH	VS	VS	VS	VS	VS	VS	VS	VS	VS	VS	Then	H
OR	If	H	-	H	H	H	H	H	H	H	H	H	H	Then	L
OR	If	V	-	V	V	V	V	V	V	V	V	V	V	Then	L
		H	-	H	H	H	H	H	H	H	H	H	H		
OR	If	M	-	M	M	M	M	M	M	M	M	M	M	Then	M

Table 5 Structure looseness rule basis for SFI (VS: very small, S: small, M: medium, H: high, VH: very high, L: low)

Logical operator		$\frac{1}{2}$x	1x	$1\frac{1}{2}$x	2x	$2\frac{1}{2}$x	3x	$3\frac{1}{2}$x	4x	$4\frac{1}{2}$x	5x	$5\frac{1}{2}$x	6x		Confidence Level
AND	If	⋯	VH	⋯	H	⋯	M	⋯	S	⋯	VS	⋯	VS	Then	H
AND	If	⋯	VH	⋯	VS	⋯	VS	⋯	VS	⋯	VS	⋯	VS	Then	L
AND	If	⋯	VH	⋯	S	⋯	S	⋯	S	⋯	S	⋯	S	Then	L
OR	If	M	⋯	M	⋯	M	⋯	M	⋯	M	⋯	M	⋯	Then	H
OR	If	H	⋯	H	⋯	H	⋯	H	⋯	H	⋯	H	⋯	Then	H
OR	If	VH	⋯	VH	⋯	VH	⋯	VH	⋯	VH	⋯	VH	⋯	Then	H
OR	If	⋯	⋯	⋯	⋯	⋯	⋯	⋯	VH	⋯	VH	⋯	VH	Then	H
AND	If	⋯	VH	⋯	VH	⋯	VH	⋯	VH	⋯	VH	⋯	VH	Then	H
AND	If	⋯	H	⋯	H	⋯	H	⋯	H	⋯	H	⋯	H	Then	H

Table 6 Misalignment rule basis for SFI (VS: very small, S: small, M: medium, H: high, VH: very high, L: low)

Logical operator		$\frac{1}{2}$x	1x	$1\frac{1}{2}$x	2x	$2\frac{1}{2}$x	3x	$3\frac{1}{2}$x	4x	$4\frac{1}{2}$x	5x	$5\frac{1}{2}$x	6x		Confidence Level
AND	If	⋯	VH	⋯	VH	⋯	VS	⋯	VS	⋯	VS	⋯	VS	Then	H
AND	If	⋯	VS	⋯	VH	⋯	VS	⋯	VS	⋯	VS	⋯	VS	Then	H
AND	If	⋯	VH	⋯	VS	⋯	VS	⋯	VS	⋯	VS	⋯	VS	Then	L
AND	If	⋯	VH	⋯	S	⋯	S	⋯	S	⋯	S	⋯	S	Then	L
OR	If	M	⋯	M	⋯	M	⋯	M	⋯	M	⋯	M	⋯	Then	L
OR	If	M	⋯	H	⋯	H	⋯	H	⋯	H	⋯	H	⋯	Then	L
OR	If	VH	⋯	VH	⋯	VH	⋯	VH	⋯	VH	⋯	VH	⋯	Then	L

are utilized by the SFI to make the decision. It is worth noting that to merely test the effectiveness of the SFI, fault-planted data were used instead of neural networks identified ones. This, however, is no shortcoming since in the SFI approach neural network detection is not directly used to obtain a decision. It rather determines the fuzzy inference rule base to be applied that is selected here according to planted fault. The fuzzy system used is the Takagi-Sugeno-Kang (TSK) model with ten inputs corresponding to each element in the normalized amplitude vector [20,21]. Each input is fuzzy partitioned with triangular membership functions.

The output is limited to singletons corresponding to three confidence levels namely, low, medium, and high. The Min and Max operators were adopted to perform fuzzy conjunctive and disjunctive operations, respectively [20]. In what follows the results for identifying three different cases will be presented. Those cases are misalignment, unbalance, and structure-looseness. The confidence levels represent the results of the series configuration.

1. **Unbalance: Main characteristics**: Large 1x, no other components. Table 4 gives the unbalance rule basis using the series configuration.
2. **Structure looseness: Main characteristics**: High 1x and harmonics, sometimes half harmonics. Table 5 gives the structure looseness rule basis using the series configuration.
3. **Misalignment: Main characteristics**: High 1x and 2x, sometimes 3x. Table 6 gives the misalignment rule basis using the series configuration.

The results of applying the series configuration are as follows:

- One looseness file was incorrectly classified as misalignment (in the parallel configuration) while the actual data indicate a looseness problem. When applying the series configuration, the neural network classifies the file as looseness and the fuzzy logic system gives only 50% confidence level for the NN result.
- Two files could not be classified using the fuzzy logic system parallel configuration. When applying the series configuration the neural networks classify the files as looseness and the fuzzy logic system gives 100% confidence level for looseness in the first case and 50% confidence level for the second case.
- The remaining files were given 100% confidence level for the NN results.

Discussion and Conclusion

This effort has clearly shown the success of neural networks, stand-alone fuzzy logic, and neuro-fuzzy (SFI) systems in diagnosing the specific faults of unbalance, looseness, and misalignment.

In particular, the tools presented here along with the experimental results illustrate the possibility of applying automated diagnosis using the neural network and fuzzy logic technologies. This is in contrast to currently available diagnostic tools based on expert systems that require human intervention and provide only suggestions to possible faults. The work presented in this paper is clearly applied in an automated manner and provides a clear and definite diagnostic result.

Each of the presented methods, neural networks alone, fuzzy logic alone, or the series neuro-fuzzy application, was quite successful in its own capacity as a diagnostic tool. In particular, the success of these technologies in diagnosis in the presence of multiple faults and at different fault levels, as described in the fault planting section, and achieving such high success rates are a clear indication of the reliability of each of these diagnostic technologies.

The work here illustrated the success of the concept using neural networks as a high resolution configuration diagnostics tool for spectral data, and the use of the fuzzy logic inference system as a low resolution configuration diagnostic tool for spectral data. Moreover, the use of hybrid neuro-fuzzy diagnostic tools has been shown to be quite successful and provides a confidence index in the diagnosed faults.

Actually, the results of the paper suggest that it is possible to devise a completely automated diagnostic system based on the series neuro-fuzzy concept, or if desired it may be possible to work with two independent diagnostic tools, one based on neural networks and the other based on fuzzy logic. In this case human intervention may be required to evaluate the results from each diagnostic method.

Finally, it is intended to extend this work to other faults, to automate the diagnostics process.

Acknowledgment

This work was funded in part by the US-Egypt Joint Board on Scientific and Technological Cooperation under Project ID Code No. OTH4-004-001, Contract/Agreement No. 110. This support is gratefully acknowledged.

References

[1] El-Shafei, A., and Rieger, N., 2003, "Automated Diagnostics of Rotating Machinery," presented at ASME Turbo Expo, Atlanta, GA, ASME Paper No. GT 2003-38453.
[2] Mobley, R. K., 1990, *An Introduction to Predictive Maintenance*, Van Nostrand Reinhold, New York.
[3] Mitchell, J. S., 1993, *An Introduction to Machinery Analysis and Monitoring*, 2nd ed., PennWell Books, Tulsa, OK.
[4] Li, C. J., and Fan, Y., 1999, "Recurrent Neural Networks for Fault Diagnosis and Severity Assessment of a Screw Compressor," ASME J. Dyn. Syst., Meas., Control, **121**, pp. 724–729.
[5] Dellomo, M. R., 1999, "Helicopter Gearbox Fault Detection: A Neural Network Based Approach," ASME J. Vibr. Acoust., **121**, pp. 265–270.
[6] David, J. S., and Babb, A. H., 1973, *Maintainability Engineering*, Pitman, New York.
[7] Wowk, V., 1991, *Machinery Vibration: Measurement and Analysis*, McGraw-Hill, New York.
[8] Randall, R. B., 1987, *Frequency Analysis*, 3rd ed., Brüel & Kjaer, Denmark.
[9] Shih-Yaug, L., and Jen-Gwo, C., 1995, "Development of a Machine Troubleshooting Expert System Via Fuzzy Multiattribute Decision-Making Approach," in *Expert Systems with Applications 81*, Pergamon, New York, pp. 187–201.
[10] Bishop, C. M., 1995, *Neural Networks for Pattern Recognition*, Oxford University Press, New York.
[11] Bezdek, J. C., 1981, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum, New York.
[12] Zurada, J. M., 1992, *Introduction to Artificial Neural Systems*," West Publishing Co., St. Paul, MN.
[13] Zadeh, L. A., 1975, "The Concept of a Linguistic Variable and its Application to Approximate Reasoning, Part 1," Inf. Sci. (N.Y.), **8**, pp. 199–249; 1975, "The Concept of a Linguistic Variable and its Application to Approximate Reasoning, Part 2," Inf. Sci. (N.Y.), **8**, pp. 301–357; 1975, "The Concept of a Linguistic Variable and its Application to Approximate Reasoning, Part 3," Inf. Sci. (N.Y.), **9**, pp. 43–80.
[14] Hassan, T. A. F., El-Shafei, A., Zeyada, Y., and Rieger, N., 2003, "Comparison of Neural Network Architectures for Machinery Fault Diagnosis," presented at ASME Turbo Expo, Atlanta, GA, ASME Paper No. GT 2003-38450.
[15] Hagan, M. T., Demuth, H. B., and Beale, M., 1996, *Neural Network Design*, PWS, Warsaw.
[16] Demuth, H., and Baele, M., 1998, "Neural Network Toolbox for Use With MATLAB," V. 3, by MathWorks, Inc.
[17] Luo, F.-A., and Unbehauen, R., 1998, *Applied Neural Networks for Signal Processing*, Cambridge University Press, Cambridge, UK.
[18] Hagan, M. T., and Menhaj, M., 1994, "Training Feedforward Networks with Marquardt Algorithm," IEEE Trans. Neural Netw., **5**(6), pp. 989–993.
[19] El-Shafei, A., 1993, "Measuring Vibration for Machinery Monitoring and Diagnostics," Shock Vib. Dig., **25**(1), pp. 3–14.
[20] Matlab Fuzzy Logic Toolbox help files, version 5.3.
[21] Sugeno, M., 1985, *Industrial Applications of Fuzzy Control*, Elsevier Science, Amsterdam.
[22] Zadeh, L. A., 1965, "Fuzzy Sets," Inf. Control., **8**, pp. 338–353.
[23] Zadeh, L. A., 1973, "Outline of a New Approach to the Analysis of Complex Systems and Decision Processes," IEEE Trans. Syst. Man Cybern., **3**(1), pp. 28–44.
[24] Zadeh, L. A., 1988, "Fuzzy Logic," Computer, **1**(4), pp. 83–93.
[25] Zadeh, L. A., 1989, "Knowledge Representation in Fuzzy Logic," IEEE Trans. Knowl. Data Eng., **1**, pp. 89–100.
[26] Jang, J.-S. R., and Sun, C.-T., 1997, *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*, Prentice-Hall, Englewood Cliffs, NJ.

Artificial Intelligence for the Diagnostics of Gas Turbines—Part I: Neural Network Approach

R. Bettocchi
M. Pinelli
P. R. Spina
M. Venturini

ENDIF Engineering Department in Ferrara,
University of Ferrara,
Via Saragat,
1-44100 Ferrara, Italy

In the paper, neural network (NN) models for gas turbine diagnostics are studied and developed. The analyses carried out are aimed at the selection of the most appropriate NN structure for gas turbine diagnostics, in terms of computational time of the NN training phase, accuracy, and robustness with respect to measurement uncertainty. In particular, feed-forward NNs with a single hidden layer trained by using a back-propagation learning algorithm are considered and tested. Moreover, multi-input/multioutput NN architectures (i.e., NNs calculating all the system outputs) are compared to multi-input/single-output NNs, each of them calculating a single output of the system. The results obtained show that NNs are sufficiently robust with respect to measurement uncertainty, if a sufficient number of training patterns are used. Moreover, multi-input/multioutput NNs trained with data corrupted with measurement errors seem to be the best compromise between the computational time required for NN training phase and the NN accuracy in performing gas turbine diagnostics. [DOI: 10.1115/1.2431391]

Introduction

In order to improve the efficiency and productivity of gas-turbine-based plants, in addition to studies aimed at the improvement of existing gas turbine technology, it is also necessary to consider the improvement and optimization of the machine management strategy during overall gas turbine operating life [1].

For an optimized machine management strategy, the study and development of suitable maintenance programs, which can ensure high levels of availability in the most efficient operating conditions, is very important [1,2]. The main gas turbine maintenance methods are (i) *emergency maintenance*, which is carried out when a fault has already appeared, (ii) *scheduled maintenance*, which is widely used in industrial applications, and (iii) *on-condition maintenance*, which, at the moment, has been widely applied only to aircraft engines [2].

Emergency maintenance should be avoided because of high costs due to damage that may be suffered by the engine and the loss of profits due to the standstill of production related to any gas turbine unscheduled stop. *Scheduled maintenance* is performed according to a priori schedules, regardless of the effective gas turbine health state. This strategy has the advantage that emergency maintenance is reduced, but this often leads to the replacement of healthy components and to gas turbine maintenance stops for a longer period than effectively necessary. The solution which seems the most suitable in order to maximize machine availability and reduce costs is to support the regular maintenance schedules with degradation demand maintenance (*on-condition maintenance*), in which maintenance actions are based on gas turbine monitoring and the actual operating state [2–9].

A review and comparison of the most effective methodologies for gas turbine monitoring and diagnostics is presented by Li [10]. A chronological list of these methodologies may include linear gas path analysis [11], linear estimators [12–15], nonlinear gas path analysis [16–19], and artificial intelligence based methods [20], such as neural networks (NNs) [21–27], fuzzy logic systems (FLSs) [28,29] and expert systems (ESs) [30–35].

Contributed by the IGTI for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received December 2, 2005; final manuscript received September 8, 2006. Review conducted by Ryohei Yokoyama. Paper presented at the ASME Turbo Expo 2005: Land, Sea and Air (GT2005), June 6, 2005–June 9, 2005, Reno, NV. Paper No. GT2005-68026.

The use of artificial-intelligence-based methods allows the overcoming of some problems related with the conventional mathematical modeling of real systems, which are highly nonlinear, time-variant systems, such as:

- the accurate assessment of the modeling approximations that can be considered acceptable.
- the lack of data, such as for gas turbines, the compressor, and turbine performance maps, which are usually unavailable.
- the computer program complexity, which may require computational times incompatible with real-time simulations.

In particular [20,36], NNs are *black-box* models capable of learning system behavior by using system input-output data. They have good generalization capabilities, so that they are usually effective in modeling nonlinear, time-variant systems, even under noisy conditions. On the contrary, due to their *black-box* nature, it is difficult to understand how the NN actually learns the input-output relationships and to determine its proper size and structure to solve a given problem.

FLSs allow the representation of uncertain and imprecise knowledge, converting complex problems into simpler problems using approximate reasoning. A FLS involves the transformation of the system crisp input values to fuzzy values (*fuzzification process*), *fuzzy inference*, which represents the *knowledge base* and allows the determination of the fuzzy outputs from the inputs, and, finally, the transformation of the fuzzy outputs to crisp output values (*defuzzification process*). FLSs allow the incorporation of the knowledge base in a simpler manner than NNs and are cost effective for a wide range of applications, compared to traditional mathematical model-based methods and, also, to NNs and traditional ESs. However, when system complexity increases, it becomes very complex (in certain cases, impossible) to find a proper set of rules and a correct tuning of the FLS to obtain a desired model accuracy.

Traditional ESs mainly manipulate symbols, on the assumption that such behavior can be stored in symbolically structured knowledge bases. However, the symbolic manipulations limit the cases to which traditional ESs can be applied, so that traditional ESs are often combined with mathematical model-based methods, NNs, or FLSs.

Among artificial-intelligence-based methods, NNs have proved to be very effective for gas turbine diagnostics [21–27] and their main advantages are the following:

- they do not require knowledge of the physics of the problem to be modeled, since they allow the reproduction of a system, once they are trained by using input-output data of the system itself.
- the capability of learning different typologies of information, such as quantitative and qualitative rules deriving from field experience.
- high robustness in the presence of poor and incorrect input data, such as the ones measured in the field on gas turbines in operation.
- high computational speed, which allows real-time calculations.

On the other hand, the main limitation of NNs is high prediction error when they operate outside the field in which they were trained, i.e. they are not able to extrapolate.

In this paper, NN models for gas turbine diagnostics are studied and developed. In particular, the data used for both training and testing the NNs were obtained by means of a cycle program, calibrated on a 255 MW single-shaft gas turbine working in the ENEL combined cycle power plant of La Spezia (Italy) [37]. In order to develop the NN model, the first step was the selection of the input and output variables and the range of variation for the inputs. Then a database containing the patterns for NN training and testing was obtained by running the cycle program with random values of the inputs, each value being within its respective range of variation. Some of the patterns were used for the NN training phase, whereas the rest were used for NN testing. In particular, the database contains variation of the *Health Indices* (which are the characteristic parameters that are indices of the gas turbine health state, such as efficiencies and characteristic flow passage areas of compressor and turbine) and the corresponding variation of the measured quantities with respect to the ones expected in *new and clean* conditions.

The analyses carried out are directed toward the selection of the most suitable NN structure for gas turbine diagnostics, in terms of computational time of the NN training phase, accuracy, and robustness toward measurement uncertainty during simulations. In particular, feed-forward NNs with a single hidden layer trained by using a back-propagation learning algorithm were considered and tested. Moreover, multi-input multioutput (MIMO) NN architectures (i.e., NNs calculating all the Health Indices) were compared to multi-input single-output (MISO) NNs, each of them calculating a single Health Index.

Gas Turbine Diagnostic Systems Based on Black-Box Models

The main problems encountered in the field application of methodologies for gas turbine monitoring and diagnostics are (i) the availability of a program for the thermodynamic cycle calculation (cycle program), which should accurately reproduce the actual machine behavior, and (ii) the available measured data that are affected by measurement errors. In particular, the first problem derives from the unavoidable modeling approximations and from the lack of accurate data for both the calibration of the cycle program on the considered gas turbine model and for the fine tuning on the particular unit.

For these reasons, the resulting errors in the estimation of the characteristic parameters of gas turbine health state (Health Indices) are usually enough high, so that, in some cases, gas turbine diagnostics can result impossible to be performed. An acceptable accuracy in the estimation of Health Indices is usually considered to be within ±1%. Nevertheless, this level of accuracy cannot be often satisfied in industrial applications, and thus, these values may be optimistic [38–42].

In order to reduce these errors, instead of calculating the values of Health Indices starting from measurements, it is advisable to evaluate the relative deviations of the Health Indices with respect to the reference values, starting form the relative deviations of the measurements with respect to the values of the measurable parameters expected in *new and clean* conditions (*baselines*). In fact, the dependence of the relative deviations of the measurable parameters on the operating point is weak and, in this manner, errors in the estimation of the relative deviations of the Health Indices can be reduced [11,42–44]. In this case, however, the main problem is the accurate estimation of the baselines [11]. The estimation of the baselines can be efficiently performed by using black-box models, such as NNs, trained with field data measured on the gas turbine in new and clean conditions.

Neural networks can be also used for gas turbine diagnostics, which consist in the evaluation of the relative deviations of the Health Indices as a function of the relative deviations of measurements, since NNs have proved robust in the presence of measurement errors [24,25]. However, the training of NNs used for this purpose requires the availability of a database that contains the relative deviations of the Health Indices as a function of the relative deviations of the measurements with respect to baselines. Since Health Indices are nonmeasurable quantities, they have to be determined by means of cycle programs [16,18].

In conclusion, a gas turbine diagnostic system can be composed of:

- a neural network for the estimation of the baselines. Such NN is trained on field data measured on the gas turbine in new and clean conditions.
- a cycle program, calibrated on the considered gas turbine, for the generation of the database containing the relative deviations of the Health Indices as a function of the relative deviations of the measurable parameters with respect to baselines.
- a neural network for gas turbine diagnostics. Such an NN is trained by using the database generated by means of the cycle program.

In particular, the paper focuses on the development and setup of NNs for gas turbine diagnostics.

Artificial Neural Networks

Artificial NNs are mathematical structures that distribute input data into several interconnected simple units (the artificial neurons) in which data are processed in parallel. Because of their high connectivity and parallelism, artificial NNs are able to link, in a nonlinear way, a multidimensional input space with a multidimensional output space, allowing very high computational speed [45].

The NN structure used in this paper is the typical *feed-forward multilayer perceptron*. In particular, NNs with one hidden layer and a continuous sigmoid activation function were used, since it has been shown that this type of NN structure is able to represent any type of multidimensional nonlinear function, if a suitable number of neurons of the hidden layer is chosen [46].

For the NN training, the back-propagation algorithm was used [47]. In the MATLAB neural network toolbox, a number of different back-propagation algorithms are present. After a preliminary analysis, the TRAINSCG algorithm was used, since it appears to be very effective and requires an acceptable computational effort. This algorithm is based on a conjugate gradient scheme in which the learning rate is varied at each iteration. The adopted stopping criterion for the NN training phase is the minimization of a performance function that was chosen to be the mean-square error (MSE) on the whole training set between the target outputs and the corresponding NN computed outputs:

Table 1 Fiat Avio 701F gas turbine main features (ISO conditions, natural gas fuel, no intake, and discharge pressure drops)

Power output	255 MW
Overall efficiency	36.0%
Exhaust mass flow rate	574 kg/s
Fuel mass flow rate	13.2 kg/s
Exhaust gas temperature	889 K
Pressure ratio	13.9
Compr. discharge temperature	655 K

$$\mathrm{MSE} = \frac{1}{n_o n_{patt}} \sum_{j=1}^{n_o} \sum_{i=1}^{n_{patt}} e_{ij}^2 = \frac{1}{n_o n_{patt}} \sum_{j=1}^{n_o} \sum_{i=1}^{n_{patt}} [t_{ij} - y_{ij}]^2 \qquad (1)$$

where n_o is the number of NN outputs, n_{patt} is the number of patterns used for the NN training, t_{ij} are the target outputs and y_{ij} the NN computed outputs. Two additional stopping criteria were used:

- attainment of an imposed maximum number of epochs (i.e., number of times that all the patterns are presented to the NN) for NN training; this number was set at 10^6.
- *early stopping method,* i.e., the training process is stopped when the *MSE* on a validation set of data increases or remains the same for an imposed number of epochs; this number was set at 10^2.

Data for NN Training and Testing

Gas Turbine Under Consideration. The gas turbine considered in the paper is a 255 MW FIAT Avio 701F single-shaft gas turbine with variable inlet guide vanes, working in the ENEL combined cycle power plant of La Spezia (Italy). For this machine, a cycle program, previously calibrated on the gas turbine under consideration, was available [37]. In Table 1, gas turbine main features are reported, while the sketch of the gas turbine model is shown in Fig. 1.

Data Generation. The data used for both training and testing the NNs were generated by means of the above-mentioned cycle program in different working points at different boundary, load, and health state conditions. Only data referring to steady-state conditions were generated, since the considered machine is a generator-drive gas turbine and it usually works in steady-state conditions. In particular, in order to take into account different gas turbine health states, the Health Indices of the gas turbine were considered as input variables for the cycle program. These parameters do not depend on the gas turbine operating point but are only sensitive to the gas turbine health state [16,18,19,25]. The characteristic parameters considered are the Health Indices related to compressor and turbine efficiencies and corrected mass flow rates (η_C, η_T, μ_C, and μ_T). According to [48], this combination of characteristic parameters represents the best solution for the number and kind of available measurements for the considered machine.

In Table 2 the Cycle Program input variables with their range of variations and the output variables are reported (column "CP I/O," I=input; O=output). A database of 50,000 random inputs was generated. Each of the values of the inputs was obtained as,

$$Q_i = (Q_i)_l + \mathrm{rand}[(Q_i)_u - (Q_i)_l], \quad \mathrm{rand} \in [0,1]$$

where $(Q_i)_u$ and $(Q_i)_l$ are the upper and lower limits of the variation range (Table 2), respectively, and "rand" is a random number uniformly distributed in the interval $[0,1]$. Owing to random number uniform distribution, Health Index variations ranging within a symmetrical interval centered at a value equal to one (±5% for efficiency and ±10% for corrected mass flow) were considered. In fact, symmetrical limits of the variation ranges allow the generation of a suitable number of patterns with all the Health Indices close to one, which represents the situation of a gas turbine in new and clean condition.

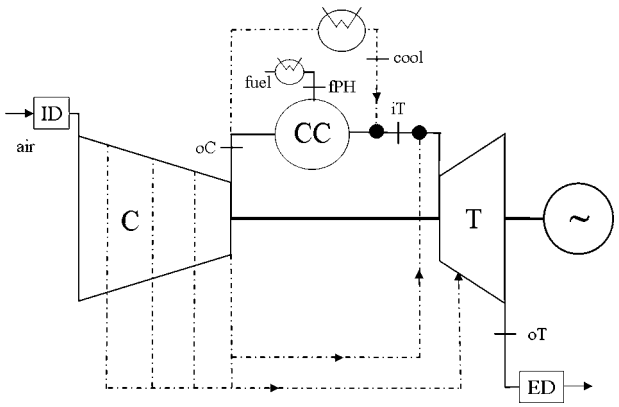


Fig. 1 FIAT Avio 701F gas turbine model

For each input set, the procedure for generating the corresponding outputs, and thus, the patterns used for NN training and testing can be outlined with the following steps:

1. The cycle program was run for the given set of input values and, in particular, for a given value of turbine inlet temperature T_{iT}. Through this calculation, the value of compressor outlet pressure p_{oC} and temperature T_{oC}, fuel mass flow rate M_f, and turbine outlet temperature T_{oT} in the current gas turbine health state condition were obtained.
2. The cycle program was run again by using the same inputs as in the previous step, with the exception of the turbine inlet temperature T_{iT}, which was replaced by the gas turbine power P_{GT} obtained in the previous calculation (step 1), and by imposing all Health Indices equal to one (new and clean condition for the machine). This allows the calculation of the expected value of p_{oC}, T_{oC}, M_f and T_{oT} in new and clean conditions.
3. The nondimensional values of the cycle program outputs p^*_{oC}, T^*_{oC}, M^*_f, and T^*_{oT} were obtained as the ratio between the value of each quantity, calculated in the current health state conditions (step 1), and the corresponding expected value in new and clean conditions (step 2), as shown in [49].

The inputs and outputs used for NN training and testing are shown in Table 2 (column "NN I/O").

Measurement Errors. In order to simulate field data and, thus, to take into account the presence of measurement errors, the generated data were corrupted with random errors included in the measurement uncertainty intervals of Table 3. The chosen measurement uncertainty values represent a situation which can be found in practice by considering a standard instrumentation category [50].

With regard to measurement uncertainty values reported in Table 3, it should be noted that:

- relative humidity uncertainty values are absolute values. Thus, since RH is expressed as a percentage, the uncertainty is also expressed as a percentage.
- the LHV uncertainty (expressed as a percentage of the reading) should be considered as an overall lack of knowledge of the LHV value, rather than a "strictly speaking" measurement uncertainty. In fact, the 3% error takes into account the fact that the on-line LHV measurement is not always performed in practice but, for example, the LHV value may be estimated from timely averaged values furnished by the fuel gas supplier.
- the uncertainties in normalized quantities (expressed as a

Table 2 Inputs (with their range of variation) and outputs of the cycle program (CP) and NN inputs and outputs

	Quantiy	Unit	Variation range	CP I/O	NN I/O
p_{amb}	Ambient pressure	kPa	[98; 103]	I	I
T_{amb}	Ambient temperature	K	[278.15; 308.15]	I	I
RH	Relative humidity	%	[20; 100]	I	I
Δp_{ID}	Pressure drop at the compressor inlet	kPa	[0; 2.5]	I	I
Δp_{ED}	Pressure drop at the turbine outlet	kPa	[0; 4.0]	I	I
N	Rotational speed	rpm	[2990; 3010]	I	I
T_{cool}	Cooling flow temperature after the exchanger	K	[453.15; 493.15]	I	I
T_{fPH}	Temperature of the fuel after heating	K	[288.15; 383.15]	I	I
LHV	Fuel lower heating value	kJ/kg	[40,000; 50,000]	I	I
A_{IGV}	IGV angular position	deg	[−4; 25]	I	I
T_{iT}	Turbine inlet temperature	K	[1150; 1750]	I	-
η_C	Compressor efficiency health index	-	[0.95; 1.05]	I	O
μ_C	Compressor corrected mass flow health index	-	[0.90; 1.10]	I	O
η_T	Turbine efficiency health index	-	[0.95; 1.05]	I	O
μ_T	Turbine corrected mass flow health index	-	[0.90; 1.10]	I	O
p^*_{oC}	Nondimensional compressor outlet pressure	-		O	I
T^*_{oC}	Nondimensional compressor outlet temperature	-		O	I
M^*_f	Nondimensional fuel mass flow rate	-		O	I
T^*_{oT}	Nondimensional turbine outlet temperature	-		O	I
P_{GT}	Gas turbine power	kW		O	I

percentage of the reading) were imposed equal to the uncertainty in the corresponding measured quantity, by considering the absolute uncertainty reported in [50].

Normalization of Data Patterns. Before the generated data are supplied to the NN, they have to be normalized to be comparable to each other. The normalization was performed with respect to the maximum and minimum value of the data set generated for each variable, i.e.,

$$(Q_i)_{norm} = \frac{Q_i - (Q_i)_{min}}{(Q_i)_{max} - (Q_i)_{min}}, \quad (Q_i)_{norm} \in [0,1]$$

A database containing 50,000 patterns of data was built; 10,000 patterns were used for the NN training phase, while the rest (40,000) were used for NN testing. In the following, the developed NNs have all been tested on the above-mentioned 40,000 patterns, not including the patterns used for training.

Influence Analysis of NN Configuration Parameters

In order to identify the most appropriate NN structure for gas turbine diagnostics, in terms of computational time of the NN training phase, accuracy, and robustness against measurement uncertainty, a sensitivity analysis was carried out with respect to the main NN configuration parameters.

The NN configuration parameters whose influence was analyzed in the paper are

- number of neurons in the hidden layer.
- optimal training set dimension, i.e., optimal number of patterns that have to be used for the NN training process.
- number of NN outputs, i.e., MIMO or MISO neural network architectures were compared.

Table 3 Measurement uncertainty values

Quantity	Meas. uncertainty	Quantity	Meas. uncertainty
p_{amb}	±0.630 [kPa]	LHV	±3.0 [%][a]
T_{amb}	±0.3 [K]	A_{IGV}	±1.0 [deg]
RH	±2.0 [%]	P_{GT}	±2.0 [%][a]
Δp_{IC}	±0.10 [kPa]	p^*_{oC}	±1.43 [%][a]
Δp_{ED}	±0.10[kPa]	T^*_{oC}	±0.23 [%][a]
N	±30.0 [rpm]	M^*_f	±1.50 [%][a]
T_{cool}	±1.0 [K]	T^*_{oT}	±0.90 [%][a]
T_{fPH}	±1.0 [K]		

[a]Percentage of the reading.

Moreover, NN robustness against measurement uncertainty was evaluated for all the considered cases.

The parameter used for the comparison of the NNs is the root-mean-square error RMSE made by the NN on the whole set of test data in the calculation of each jth output ($\eta_C, \mu_C, \eta_T, \mu_T$)

$$\left(\text{RMSE} = \sqrt{\frac{1}{n_{patt}} \sum_{i=1}^{n_{patt}} \left(\frac{t_i - y_i}{t_i} \right)^2} \right)_j, \quad j = 1, \ldots, n_o \qquad (2)$$

where t_i are the target outputs, y_i the computed outputs, n_{patt} is the number of patterns used for NN testing, and no is the number of NN outputs.

Number of Neurons in the Hidden Layer. The influence of the number of neurons in the hidden layer on the NN simulation error was evaluated by comparing the response of different NNs characterized by different numbers of neurons in the hidden layer. Each NN was trained by using 8000 patterns, taken from the 10,000 available training patterns, while the RMSE was evaluated by running each NN in order to reproduce 40,000 patterns, not including the ones used for training. All patterns used for both training and testing were not affected by measurement errors.

The analysis of the influence of the number of neurons in the hidden layer substantially confirmed the results presented in [50], i.e., the number of neurons in the hidden layer should be higher than twice the number of inputs. In particular, 60 neurons in the hidden layer seems to be a good compromise between NN accuracy and computational time required for the NN training phase. In the analyses carried out in the rest of the paper, only NNs with 60 neurons in the hidden layer were used.

Number of Training Patterns and NN Architecture. The analyses were focused on the evaluation of the influence of:

- number of patterns used for the training phase; in particular, 1000, 2000, 4000, and 8000 patterns were used for training. In fact, for a number of training patterns lower than 1000, it was verified that errors can be very high, due to the *overfitting* phenomenon, even if the *early stopping method* is used.

- NN architecture; multi-input/multioutput (MIMO) or multi-input/single-output (MISO) architectures were analyzed;
- presence of measurement errors; two situations were investigated:
  - presence of measurement errors only in the test patterns, i.e., training patterns are not corrupted with measurement errors, while test patterns are corrupted with measurement errors (case Tr U, Ts C).
  - presence of measurement errors both in training and test patterns, i.e., both training and test patterns are corrupted with measurement errors (case Tr C, Ts C).

In the test phase, the RMSE was evaluated by running the NN in order to reproduce the 40,000 test patterns, which, as usual, do not include the patterns used for training.

The results are reported in Figs. 2(*a*)–2(*d*), for 1000, 2000, 4000, and 8000 training patterns, respectively.

For comparison purposes, the case in which both training and test patterns are not corrupted with measurement errors (case Tr U, Ts U) was also reported. In fact, as explained in [50], the RMSE value calculated in this case accounts for all the errors made in the calculation of predicted values (simulation numerical solution uncertainty, simulation modeling uncertainty due to the use of the selected training data and simulation modeling uncertainty due to modeling assumptions, such as NN structure, number of hidden layers and of neurons of each hidden layer, training algorithm, etc.). Thus, this RMSE value represents the error, which can be attributed to the NN model only, while it is not influenced by measurement errors.

In the case in which both training and test patterns are not corrupted with measurement errors (case Tr U, Ts U), RMSE values are clearly acceptable, since the highest RMSE is $\sim 0.7\%$ (corrected mass flows μ_C estimated through a MIMO NN trained with 1000 training patterns, Fig. 2(*a*)).

With regard to the cases in which test data are corrupted by measurement errors, it can be noted that:

- in the case of training data not corrupted by measurement errors (case Tr U, Ts C), RMSE remains almost constant, by varying the number of training patterns. Otherwise, in the case of training data corrupted by measurement errors (case Tr C, Ts C), RMSE decreases as the number of training patterns increases. Since for 1000 training patterns the RMSE in case (Tr C, Ts C) is higher than RMSE in case (Tr U, Ts C), the use of training data corrupted by measurement errors becomes convenient only for a number of training patterns equal or higher than 4000. In any case, the benefit of increasing the number of training patterns reduces for high numbers of training patterns, so that it becomes negligible when 8000 training patterns (Fig. 2(*d*)) are used instead of 4000 (Fig. 2(*c*)), as shown in [50].
- for NNs trained by using data not corrupted with measurement errors, MISO networks reach lower values of RMSE than the corresponding MIMO network, whereas the contrary occurs for networks trained by using data corrupted with measurement errors. In any case, the difference between MIMO and MISO networks reduces by augmenting the number of training patterns.
- RMSE values on efficiencies η_C and η_T are lower than RMSE on corrected mass flows μ_C and μ_T. RMSE values always lower than 1% are reached for η_C and η_T, by using only 2000 training patterns (Fig. 2(*b*)), whereas RMSE values for μ_C and μ_T remain in the range 1.5–2.0% even if 8000 training patterns are used (Fig. 2(*d*)).

The evaluation of the error distribution showed that it is random in the whole range of variation of NN input variables. A detailed analysis for the case of the MIMO NN trained with 4000 patterns corrupted with measurement errors is reported in the Appendix.

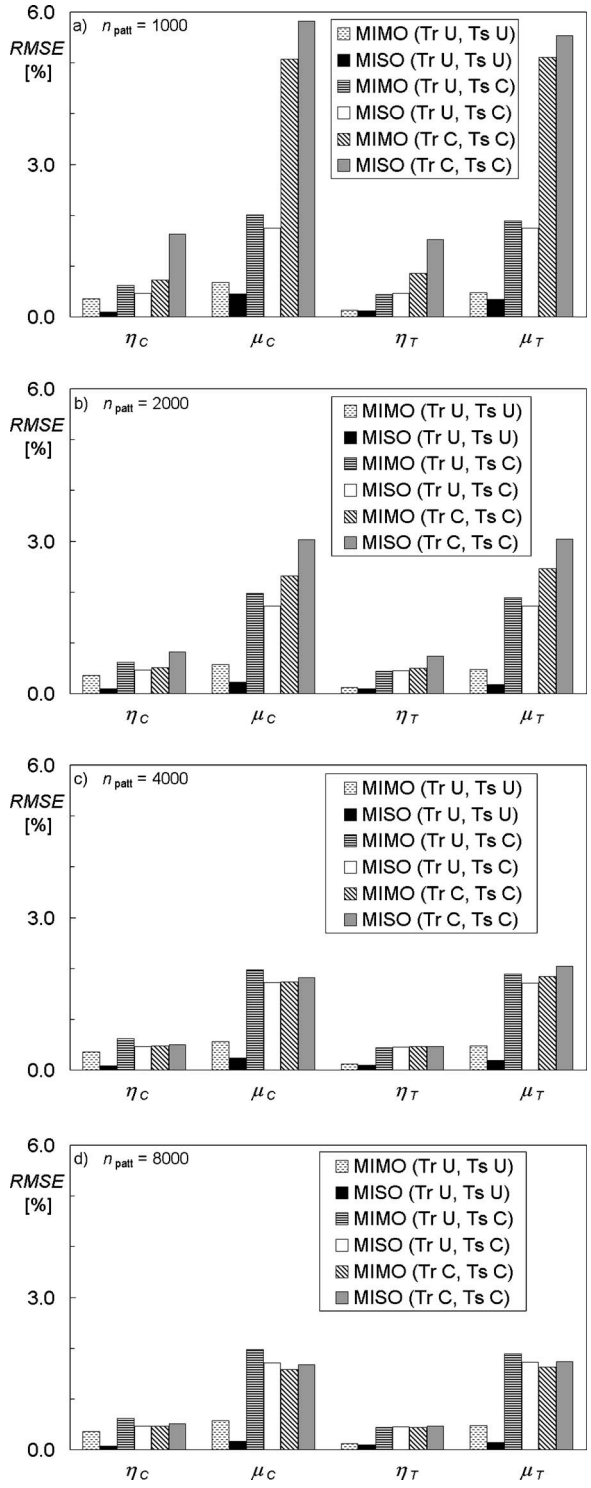


Fig. 2 Influence of the number of outputs (MIMO and MISO) in the presence of measurement uncertainty—Number of training patterns: (*a*) 1000, (*b*) 2000, (*c*) 4000, and (*d*) 8000

Validation of NN Models. For the validation of the NN models, it was considered that a success in the simulation of a computed output y_j is achieved when it results that

$$|e|_j = |t - y|_j \leq \delta, \quad j = 1, \ldots, n_o \tag{3}$$

where δ is a chosen threshold. Four different thresholds have been considered: 0.5%, 1.0%, 2.0%, and 4.0%. This allows the calculation of the success rate (SR) for each NN output, defined as the

Table 4 Success rate for MISO and MIMO networks (4000 training patterns) SR=100% (set in bold)

Threshold δ [%]		Success rate SR[%] 0.5			1.0			2.0		4.0			
	Training	U	U	C	U	U	C	U	U	C	U	U	C
	Testing	U	C	C	U	C	C	U	C	C	U	C	C
η_C	MISO	**100**	69	66	**100**	98	96	**100**	**100**	**100**	**100**	**100**	**100**
	MIMO	93	55	68	**100**	89	97	**100**	**100**	**100**	**100**	**100**	**100**
μ_C	MISO	94	22	22	99	42	42	**100**	73	73	**100**	99	98
	MIMO	36	19	23	99	37	44	**100**	67	75	**100**	96	99
η_T	MISO	**100**	71	71	**100**	98	97	**100**	**100**	**100**	**100**	**100**	**100**
	MIMO	**100**	72	71	**100**	98	97	**100**	**100**	**100**	**100**	**100**	**100**
μ_T	MISO	96	22	21	**100**	43	40	**100**	74	70	**100**	99	96
	MIMO	70	20	22	99	39	43	**100**	69	74	**100**	97	98

percentage ratio between the number of successes and the number of test patterns n_{patt}.

The SR values obtained for the case of NNs trained by using 4000 training patterns, with both MIMO and MISO architectures, are reported in Table 4, where, for each considered threshold, the results for NNs trained with data not corrupted with measurement errors, but tested on data that are corrupted with measurement errors (Tr U, Ts C), are compared to the ones obtained by using NNs both trained and tested with data corrupted with measurement errors (Tr C, Ts C).

For comparison purposes, SR values in the case in which both training and test patterns are not corrupted with measurement errors (case Tr U, Ts U) was also reported, as done while comparing RMSE values for the different cases. In this last case, SR values are very high. In particular, for MISO models, all outputs can be calculated with an error lower than 0.5% in more than 94% of the cases.

With regard to the cases in which test data are corrupted by measurement errors, it can be noted that:

- when data not affected by measurement errors are used for training, SR values for MISO NNs are usually higher than the ones obtained for MIMO NNs; the contrary occurs when data affected by measurement errors are used for training.
- the comparison between the best MIMO and MISO NNs (for any given value of the threshold δ and for any considered NN output) shows that SR values are comparable, i.e., there is no appreciable difference between the two NN architectures. Thus, since the computational time required for the training phase of a single MISO NN is comparable to that required for a MIMO NN (i.e., for four MISO NNs, an overall training time about four times the time necessary for training a MIMO NN is required), the best solution seems to be the MIMO NN trained with data corrupted with measurement errors.
- the SR obviously increases as the value of the threshold δ increases. In any case, as can be observed from Fig. 3, reporting the success rates for the MIMO NN trained with 4000 patterns corrupted with measurement errors, results in
  - SRs for efficiencies η_C and η_T are always higher than those for corrected mass flows μ_C and μ_T.
  - errors on efficiencies η_C and η_T are very small ($<1\%$ in 97% of cases); whereas for corrected mass flows μ_C and μ_T, errors lower than 2% occur in about 75% of cases.

These results are in agreement with the ones reported in Fig. 2(*c*) by considering that lower RMSE values imply higher values for the success rate.

Conclusions

In this paper, neural network (NN) models for gas turbine diagnostics have been studied and developed. The considered machine was a 255 MW single-shaft gas turbine, whose data were generated for different boundary, load, and health state conditions by means of a cycle program. The analyses carried out were focused on the selection of the most appropriate NN structure for gas turbine diagnostics, in terms of computational time of the NN training phase, accuracy, and robustness with respect to measurement uncertainty.

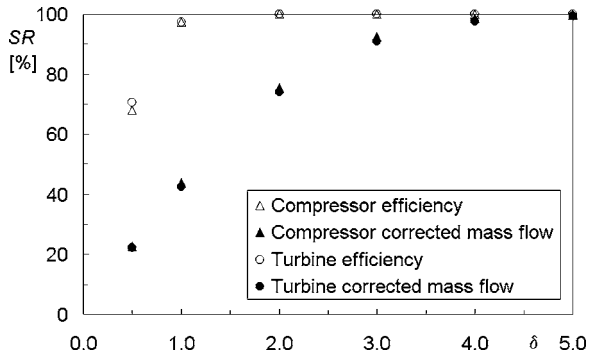


Fig. 3 Success rate versus threshold values δ for MIMO NN trained by using 4000 patterns corrupted with measurement errors

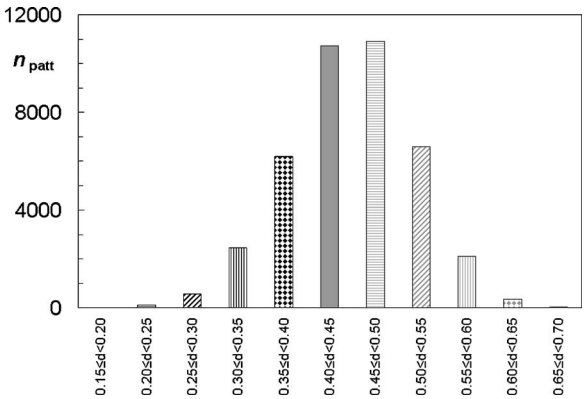


Fig. 4 Number of generated patterns versus normalized Euclidean distance *d*

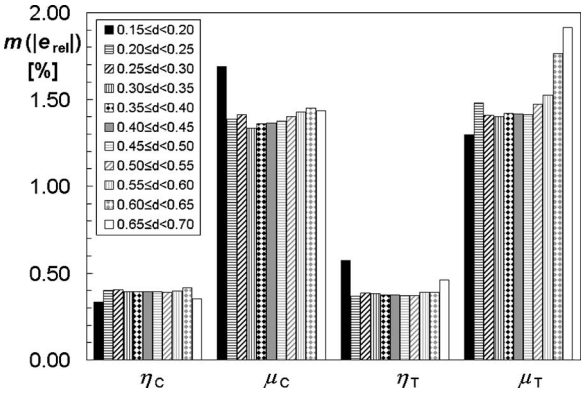


Fig. 5 Mean value of the absolute relative error for different intervals of the normalized Euclidean distance *d*

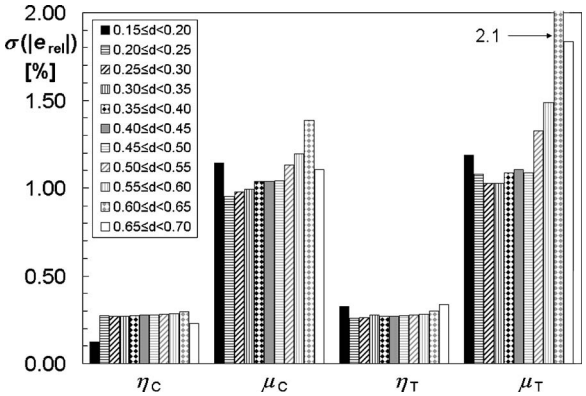


Fig. 7 Standard deviation of the absolute relative error for different intervals of the normalized Euclidean distance *d*

For the considered 15 inputs/4 outputs system, the most appropriate NN structure proved to be a feed-forward, multilayer perceptron with a single 60-neuron hidden layer. In the case in which test data are corrupted by measurement errors, it was observed that it is convenient to use training data not corrupted by measurement errors if the number of training patterns is lower than 4000, while lower RMSE values can be obtained by using 4000 or more training patterns corrupted by measurement errors. Moreover, no appreciable difference between the performances of MIMO and MISO neural network architectures was highlighted. Finally, the solution that represents the best compromise between the computational time required for NN training phase and the NN accuracy in performing gas turbine diagnostics is a MIMO neural network trained with 4000 patterns corrupted with measurement errors. In this case, errors on efficiency Health Indices are very small (lower than 1% in 97% of cases); whereas, for corrected mass flow Health Indices, errors lower than 2% occur in ~75% of cases.

Some general conclusions for an optimal setup of neural network models can be drawn:

- the analysis of both RMSE and success rate values shows that the accuracy of NN model can be considered acceptable, since high errors and low success rate values are mainly due to the presence of measurement errors.
- for test data corrupted by measurement errors, it is convenient to use training data not corrupted by measurement errors if the number of training patterns is low. If a high number of training patterns is available, it is possible to further reduce the RMSE values by using training patterns corrupted by measurement errors.

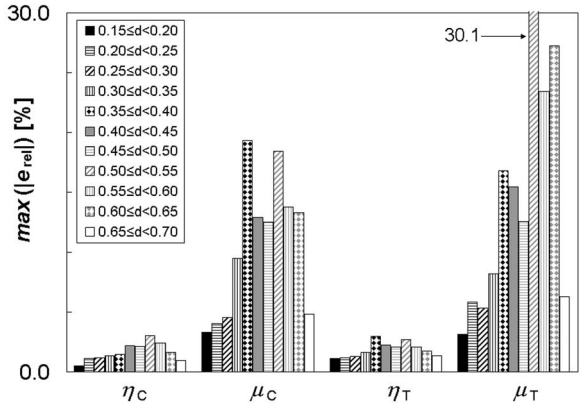


Fig. 6 Maximum absolute relative error for different intervals of the normalized Euclidean distance *d*

- for data affected by measurement errors, the number of neurons in the hidden layer slightly affects NN performance.
- since there is no appreciable difference between the performances of MIMO and MISO neural networks and since the computational time required to train a single MISO NN is comparable to that required for a MIMO NN, it is usually preferable to adopt a MIMO architecture.

Thus, neural networks represent a quite easy-to-implement solution both for modeling the behavior and for performing the diagnostics of gas turbines.

Acknowledgment

The work was carried out with the support of the MiUR (Italian Ministry of University and Scientific & Technological Research).

Nomenclature

A_{IGV} = inlet guide vane angular position
C = corrupted (with measurement error)
CP = cycle program
e = $=t-y$ error
ES = expert system
FLS = fuzzy logic system
I = input
LHV = fuel lower heating value
M = mass flow rate
MIMO = multi-input/multioutput
MISO = multi-input/single-output
MSE = mean square error
n_o = number of outputs
n_{patt} = number of patterns
N = rotational speed
NN = neural network
O = output
p = pressure
P = power
Q = quantity
rand = uniformly distributed random number
RH = relative humidity
RMSE = root-mean-square error
SR = success rate
t = expected target output
T = temperature
Tr = training data
Ts = testing data
U = uncorrupted (without measurement error)
y = computed output
δ = threshold
Δ = variation

η = efficiency health index
μ = corrected mass flow rate health index

Subscripts and Superscripts

* = nondimensional value
amb = ambient
cool = cooling flow after the heat exchanger
C = compressor
ED = exhaust duct
f = fuel
GT = gas turbine
i = inlet section
ID = inlet duct
l = lower limit
max = maximum
min = minimum
norm = normalized value
o = outlet section
PH = postheating
T = turbine
u = upper limit

Appendix

The output calculated by the NN is a function of the value assumed by the vector $\boldsymbol{x}=[x_1,\ldots,x_n]^{\mathrm{T}}$ of NN inputs (n=number of NN inputs). Each vector $\boldsymbol{x}$ defines a gas turbine operating point, which can be identified through its Euclidean distance from a machine reference operating point.

A normalized Euclidean distance for each pattern used for NN testing (n_{patt}=40,000) was defined as

$$d_i=\left[\sqrt{\frac{1}{n}\sum_{k=1}^{n}\frac{(x_k-x_{d,k})^2}{(x_k-x_{d,k})^2_{\max}}}\right]_i,\quad d_i\in[0,1]\quad i=1,\ldots,n_{\mathrm{patt}} \tag{A1}$$

where the vector $\boldsymbol{x}_d=[x_{d,1},\ldots,x_{d,n}]^{\mathrm{T}}$ identifies the machine reference operating point (assumed equal to the design point) and the term $|x_k-x_{d,k}|_{\max}$ represents, for each input, the maximum distance from the machine reference point.

In order to evaluate the error distribution for each of the four NN outputs, the absolute relative error

$$|e_{\mathrm{rel}}|_{ij}=\left|\frac{t_{ij}-y_{ij}}{t_{ij}}\right|,\quad i=1,\ldots,n_{\mathrm{patt}};\ j=1,\ldots,n_{\mathrm{o}} \tag{A2}$$

was calculated as a function of the distance d_i.

The results presented in the following refer to the case of the MIMO NN trained with 4000 patterns corrupted with measurement errors, but they are representative of all the analyzed cases. In Fig. 4, the number of generated patterns for various intervals (width equal to 0.05) of the normalized Euclidean distance are reported. It can be noted that all the generated patterns lie in the range [0.15; 0.70] of the normalized Euclidean distance and that the trend of the number of generated patterns versus the normalized Euclidean distance is approximately Gaussian. This can be attributed to the fact that the number of data patterns (n_{patt} =40,000), generated by using random values of each input uniformly distributed in the respective range of variation (reported in Table 2), is finite.

The mean value $m(|e_{\mathrm{rel}}|)$ and the maximum value $\max(|e_{\mathrm{rel}}|)$ of the absolute relative error for the different intervals of the normalized Euclidean distance are reported in Figs. 5 and 6, respectively, and the standard deviation of the absolute relative error with respect to the mean value $m(|e_{\mathrm{rel}}|)$ is reported in Fig. 7. It can be observed that (i) the dependence of the mean value of the absolute relative error from the normalized Euclidean distance is weak and, as expected, (ii) the highest errors usually fall in correspondence of the normalized Euclidean distance intervals for which the data pattern density is high, while the highest standard deviations usually fall where the data pattern density is low. As a conclusion, it can be stated that the error distribution is random with respect to the considered operating point and this confirms the validity of the analyses developed throughout the paper.

References

[1] Hoeft, R. F., 1996, "Heavy Duty Gas Turbine Operating & Maintenance Considerations," *Proc. of 39th GE Turbine State-of-the-Art Technology Seminar*, GE ed., Schenectady, NY, August 26–29, Paper No. GER-3620D.
[2] Rajamani, R., Wang, J., and Jeong, K.-Y., 2004, "Condition-Based Maintenance for Aircraft Engines," ASME Paper No. GT2004-54127.
[3] Meher-Homji, B. C., and Cullen, P. J., 1992, "Integration of Condition Monitoring Technologies for the Health Monitoring of Gas Turbines," ASME Paper No. 92-GT-52.
[4] Meher-Homji, B. C., Boyce, P. M., Lakshminarashima, A. N., Whitten, A. J., and Meher-Homji, J. F., 1993, "Condition Monitoring and Diagnostic Approaches for Advanced Gas Turbines," *Proc. of 7th ASME COGEN-TURBO*, Bournemouth, UK, IGTI Vol. 8, pp. 347–354.
[5] Madej, J., Longtin, K., and Smith, D. P., 1996, "Monitoring Service Delivery System and Diagnostics," *Proc. of 39th GE Turbine State-of-the-Art Technology Seminar*, GE ed., Schenectady, NY, August 26–29, Paper No. GER-3956.
[6] Tsalavoutas, A., Aretakis, N., Mathioudakis, K., and Stamatis, A., 2000, "Combining Advanced Data Analysis Methods for the Constitution of an Integrated Gas Turbine Condition Monitoring and Diagnostic System," ASME Paper No. 2000-GT-0034.
[7] Bettocchi, R., Pinelli, M., Spina, P. R., Venturini, M., and Sebastanelli, S., 2001, "A System for Health State Determination of Natural Gas Compression Gas Turbines," ASME Paper No. 2001-GT-0223.
[8] Veer, T., Ulvestad, A., and Bolland, O., 2004, "Frame, a Toll for Predicting Gas Turbine Condition as Well as Reliability, Availability Performance," ASME Paper No. GT2004-53770.
[9] Byington, C. S., Roemer, M. J., Watson, M. J., Galie, T. R., and Savage, C., 2004, "Prognostic Enhancement to Diagnostic Systems (PEDS) Applied to Shipboard Power Generation Systems," ASME Paper No. GT2004-54135.
[10] Li, Y. G., 2002, "Performance Analysis Based Gas Turbine Diagnostics: A Review," Proc. Inst. Mech. Eng., Part A, **216**, pp. 363–377.
[11] Urban, L. A., 1972, "Gas Path Analysis Applied to Turbine Engine Condition Monitoring," *Proc. of AIAA/SAE 8th Joint Propulsion Conference*, New Orleans, AIAA, Washigton, DC, AIAA Paper No. 72-1082.
[12] Stamatis, A., and Papailiou, K. D., 1988, "Discrete Operating Conditions Gas Path Analysis," *Proceedings, AGARD Conference*, Quebec, Canada, pp. 33-1–33-10.
[13] Urban, L. A., and Volponi, A. J., 1992, "Mathematical Methods of Relative Engine Performance Diagnostics," SAE Trans., **101**, p. 922048.
[14] Doel, D. L., 1994, "TEMPER—A Gas-Path Analysis Tool for Commercial Jet Engines," ASME J. Eng. Gas Turbines Power, **116**, pp. 82–89.
[15] Doel, D. L., 1994, "An Assessment of Weighted-Least-Squares Based Gas Path Analysis," ASME J. Eng. Gas Turbines Power, **116**, pp. 366–373.
[16] Stamatis, A., Mathioudakis, K., and Papailiou, K. D., 1990, "Adaptive Simulation of Gas Turbine Performance," ASME J. Eng. Gas Turbines Power, **112**, pp. 168–175.
[17] Benvenuti, E., Bettocchi, R., Cantore, G., Negri di Montenegro, G., and Spina, P. R., 1993, "Gas Turbine Cycle Modeling Oriented to Component Performance Evaluation from Limited Design or Test Data," *Proc. of 7th ASME COGEN-TURBO*, Bournemouth, UK, ASME, New York, IGTI Vol. 8, pp. 327–337.
[18] Bettocchi, R., and Spina, P. R., 1999, "Diagnosis of Gas Turbine Operating Conditions by Means of the Inverse Cycle Calculation," ASME Paper No. 99-GT-185.
[19] Pinelli, M., Spina, P. R., and Venturini, M., 2003, "Optimized Operating Point Selection for Gas Turbine Health State Analysis by Using a Multi-Point Technique," ASME Paper No. GT2003-38191.
[20] Jang, J.-S. R., Sun, C.-T., and Mizutani, E., 1997, *Neuro-Fuzzy and Soft Computing*, Prentice-Hall, Englewood Cliffs, NJ.
[21] Torella, G., and Lombardo, G., 1996, "Neural Networks for the Diagnostics of Gas Turbine Engines," ASME Paper No. 96-TA-39.
[22] Kanelopoulos, K., Stamatis, A., and Mathioudakis, K., 1997, "Incorporating Neural Networks Into Gas Turbine Performance Diagnostics," ASME Paper No. 97-GT-35.
[23] Volponi, A. J., DePold, H. R., Ganguli, R., and Daguang, C., 2000, "The Use of Kalman Filter and Neural Networks Methodologies in Gas Turbine Performance Diagnostics: A Comparative Study," ASME Paper No. 2000-GT-0547.
[24] Romessis, C., Stamatis, A., and Mathioudakis, K., 2001, "A Parametric Investigation of the Diagnostic Ability of Probabilistic Neural Networks on Turbofan Engines," ASME Paper No. 2001-GT-0011.
[25] Bettocchi, R., Spina, P. R., and Torella, G., 2002, "Gas Turbine Health Indices Determination by Using Neural Networks," ASME Paper No. GT-2002-30276.
[26] Arriagada, J., Genrup, M., Loberg, A., and Assadi, M., 2003, "Fault Diagnosis System for an Industrial Gas Turbine by Means of Neural Networks," *Proc. International Gas Turbine Congress 2003 (IGTC'03)*, Tokyo, Nov. 2–7, GTSJ, Tokyo, Paper No. IGTC2003, Tokyo TS-001.
[27] Sampath, S., and Singh, R., 2004, "An Integrated Fault Diagnostics Model Using Genetic Algorithm and Neural Networks," ASME Paper No. GT-2004-53914.
[28] Ganguli, R., 2003, "Application of Fuzzy Logic for Fault Isolation of Jet

Engines," ASME J. Eng. Gas Turbines Power, **125**, pp. 617–623.
[29] Ganguli, R., Verma, R., and Roy, N., 2004, "Soft Computing Application for Gas path Fault Isolation," ASME Paper No. GT-2004-53209.
[30] Console, L., and Torasso, P., 1989, *Diagnostic Problem Solving: Combining Heuristic, Approximate and Causal Reasoning*, Van Nostrand Reinhold, New York.
[31] Doel, D. L., 1990, "The Role of Expert Systems in Commercial Gas Turbine Engine Monitoring," ASME Paper No. 90-GT-374.
[32] Torella, G., 1992, "Expert Systems for the Trouble-Shooting and the Diagnostics of Engines," AIAA Paper No. 92-3327.
[33] Palmer, C. A., 1998, "Combining Bayesian Belief Networks With Gas Path Analysis for Test Cell Diagnostics and Overhaul," ASME Paper No. 98-GT-168.
[34] DePold, H. R., and Gass, F. D., 1999, "The Application of Expert Systems and Neural Networks to Gas Turbine Prognostics and Diagnostics," ASME J. Eng. Gas Turbines Power, **121**, pp. 607–612.
[35] Spina, P. R., Torella, G., and Venturini, M., 2002, "The Use of Expert Systems for Gas Turbine Diagnostics and Maintenance," ASME Paper No. GT-2002-30033.
[36] Jain, L. C., and Martin, N. M., 1998, *Fusion of Neural Networks, Fuzzy Sets, and Genetic Algorithms*, CRC Press, Boca Raton.
[37] Bettocchi, R., Pinelli, M., Venturini, M., Spina, P. R., Bellagamba, S., and Tirone, G., 2002, "Procedura di Calibrazione del Programma per la Diagnosi Funzionale dei Turbogas della Centrale a Ciclo Combinato di La Spezia," *Proc. 57th Congresso Nazionale ATI*, Pisa, Italy, ETS ed., September 17–20, pp. III-B-3–III-B-12 (in Italian).
[38] Stamatis, A., Mathioudakis, K., and Papailiou, K., 1992, "Optimal Measurement and Health Index Selection for Gas Turbine Performance Status and Fault Diagnosis," ASME J. Eng. Gas Turbines Power, **114**, pp. 209–216.
[39] Bettocchi, R., Spina, P. R., and Benvenuti, E., 2000, "Set-Up of an Adaptive Method for the Diagnosis of Gas Turbine Operating State by Using Test-Bench Measurements," ASME Paper No. 2000-GT-0309.
[40] Mathioudakis, K., and Tsalavoutas, A., 2001, "Uncertainty Reduction in Gas Turbine Performance Diagnostics by Accounting for Humidity Effects," ASME Paper No. 2001-GT-0010.
[41] Pinelli, M., and Spina, P. R., 2002, "Gas Turbine Field Performance Determination: Sources of Uncertainties," ASME J. Eng. Gas Turbines Power, **124**, pp. 155–160.
[42] Mathioudakis, K., 2003, "Non-linear Methods for Gas Turbine Fault Diagnostics," Von Karman Institute Lecture Series 2003-01, Gas Turbine Monitoring & Fault Diagnosis, Jan. 13–17.
[43] Mathioudakis, K., Kamboukos, Ph., and Stamatis, A., 2002, "Turbofan Performance Deterioration Tracking Using Non-linear Models and Optimization Techniques," ASME Paper No. GT-2002-30026.
[44] Mathioudakis, K., and Kamboukos, Ph., 2004, "Assessment of the Effectiveness of Gas Path Diagnostic Schemes," ASME Paper No. GT-2004-53862.
[45] Haykin, S., 1999, *Neural Networks—A Comprehensive Foundation*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ.
[46] Cybenko, G., 1989, "Approximation by Superimposition of a Sigmoidal Function," Math. Control, Signals, Syst., **2**, pp. 303–314.
[47] Rumenlhart, D. E., Hinton, G. E., and Williams, R. J., 1986, "Learning Internal Representation by Error Propagation, Parallel Distributed Processing," Explor. Microstruct. Cognition, **1**, pp. 318–362.
[48] Bettocchi, R., Pinelli, M., Spina, P. R., and Venturini, M., 2003, Statistical Analyses to Improve Gas Turbine Diagnostics Reliability, *Proc. of 8th International Gas Turbine Congress*, 2003 (IGTC'03), Tokyo, November 2–7 GTSJ, Tokyo, Paper No. IGTC2003Tokyo TS-004.
[49] Pinelli, M., and Venturini, M., 2002, "Application of Methodologies to Evaluate the Health State of Gas Turbines in a Cogenerative Combined Power Plant," ASME Paper No. GT-2002-30248.
[50] Bettocchi, R., Pinelli, M., Spina, P. R., Venturini, M., and Burgio, M., 2004, "Set Up of a Robust Neural Network for Gas Turbine Simulation," ASME Paper No. GT2004-53421.

Artificial Intelligence for the Diagnostics of Gas Turbines—Part II: Neuro-Fuzzy Approach

R. Bettocchi
M. Pinelli
P. R. Spina
M. Venturini

ENDIF Engineering Department in Ferrara,
University of Ferrara,
Via Saragat,
1-44100 Ferrara, Italy

*In the paper, neuro-fuzzy systems (NFSs) for gas turbine diagnostics are studied and developed. The same procedure used previously for the setup of neural network (NN) models (Bettocchi, R., Pinelli, M., Spina, P. R., and Venturini, M., 2007, ASME J. Eng. Gas Turbines Power, **129**(3), pp. 711–719) was used. In particular, the same database of patterns was used for both training and testing the NFSs. This database was obtained by running a cycle program, calibrated on a* 255 MW *single-shaft gas turbine working in the ENEL combined cycle power plant of La Spezia (Italy). The database contains the variations of the Health Indices (which are the characteristic parameters that are indices of gas turbine health state, such as efficiencies and characteristic flow passage areas of compressor and turbine) and the corresponding variations of the measured quantities with respect to the values in new and clean conditions. The analyses carried out are aimed at the selection of the most appropriate NFS structure for gas turbine diagnostics, in terms of computational time of the NFS training phase, accuracy, and robustness towards measurement uncertainty during simulations. In particular, adaptive neuro-fuzzy inference system (ANFIS) architectures were considered and tested, and their performance was compared to that obtainable by using the NN models. An analysis was also performed in order to identify the most significant ANFIS inputs. The results obtained show that ANFISs are robust with respect to measurement uncertainty, and, in all the cases analyzed, the performance (in terms of accuracy during simulations and time spent for the training phase) proved to be better than that obtainable by multi-input/multioutput (MIMO) and multi-input/single-output (MISO) neural networks trained and tested on the same data.* [DOI: 10.1115/1.2431392]

Introduction

Artificial intelligence-based methods (such as neural networks (NNs), fuzzy logic systems (FLSs), expert systems, genetic algorithms, probabilistic reasoning, etc.) have received growing attention in recent years in solving complex nonlinear, time-variant problems [1,2]. NNs and FLSs have proved to be very effective in solving many real problems, such as, for example, diagnostic problems and, in particular, gas turbine diagnostics [3–12]. However, both methods present some limitations [2].

NNs are *black-box* models capable of learning the behavior of a system by using system input-output data. Once they are trained, NNs are able to simulate the system with a prespecified accuracy. They have good generalization capabilities, so that they are usually effective in modeling nonlinear, time-variant systems, even under noisy conditions. On the contrary, due to their black-box nature, it is difficult to understand how the NN actually learns the input-output relationships and to determine its optimal size and structure in order to solve a given problem. Moreover, NN implementation is usually more time consuming than other technologies, such as, in particular, FLSs.

FLSs allow the representation of uncertain and imprecise knowledge, converting complex problems into simpler problems using approximate reasoning. A FLS involves (i) the transformation of the system crisp input values to fuzzy values (*fuzzification process*), (ii) *fuzzy inference*, which represents the *knowledge base* and allows the determination of the fuzzy outputs from the inputs, and, finally, (iii) the transformation of the fuzzy outputs to crisp output values (*defuzzification process*). FLSs allow the incorporation of the knowledge base in a simpler manner than NNs and are cost-effective for a wide range of applications compared to traditional mathematical model-based methods, and, also, to NNs. However, when system complexity increases, it becomes very complex (in certain cases impossible) to find a proper set of rules and a correct tuning of the FLS to obtain a good solution. The generalization capability of FLSs is poor if compared to NNs, and once the rules are determined, they remain fixed, so that the FLS is unable to learn. Moreover, FLSs cannot generate rules that are able to obtain a prespecified accuracy of FLS outputs. In fact, accuracy can only be improved by a trial-and-error procedure.

Thus, an appropriate combination of these two techniques (NNs and FLSs) can effectively solve the problems of NNs and FLSs [2]. NN techniques can be used to learn system behavior from system input-output data. Then, the learned knowledge can be used to automatically generate the FLS, so reducing development time significantly.

The systems obtained by combining NNs and FLSs are called neuro-fuzzy systems (NFSs). They are able to generate FLSs of complex systems for which a conventional fuzzy approach may fail. Moreover, NFSs are *gray-box* models, since, in contrast to NNs, it is possible to operate directly on the NFS structure, for example for tuning NFS output accuracy.

In this paper, NFSs for gas turbine diagnostics are studied and developed. The same procedure previously used for the setup of NN models [12] was used. In particular, the same database of patterns was used for both training and testing the NFSs. This database was obtained by running a cycle program, calibrated on a 255 MW single-shaft gas turbine working in the ENEL combined cycle power plant of La Spezia (Italy) [13]. The database contains the variations of the *Health Indices* (which are the characteristic parameters that are indices of gas turbine health state, such as efficiencies and characteristic flow passage areas of com-

Contributed by the IGTI of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received December 2, 2005; final manuscript received September 8, 2006. Review conducted by Ryohei Yokoyama. Paper presented at the ASME Turbo Expo 2005: Land, Sea and Air (GT2005), June 6–9, 2005, Reno, NV. Paper Number GT2005-68027.

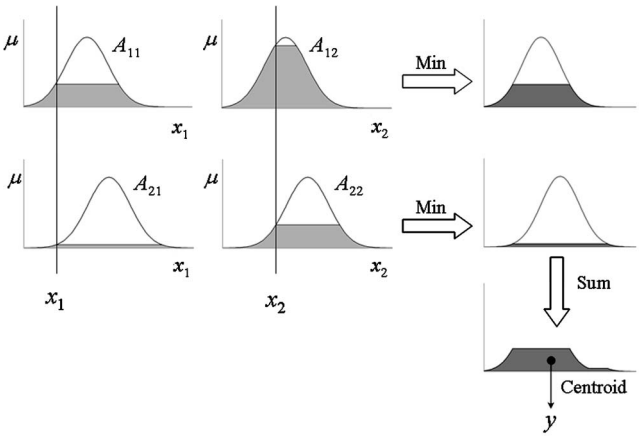


Fig. 1 Two-crisp-input Mamdani FLS with two MFs for each input

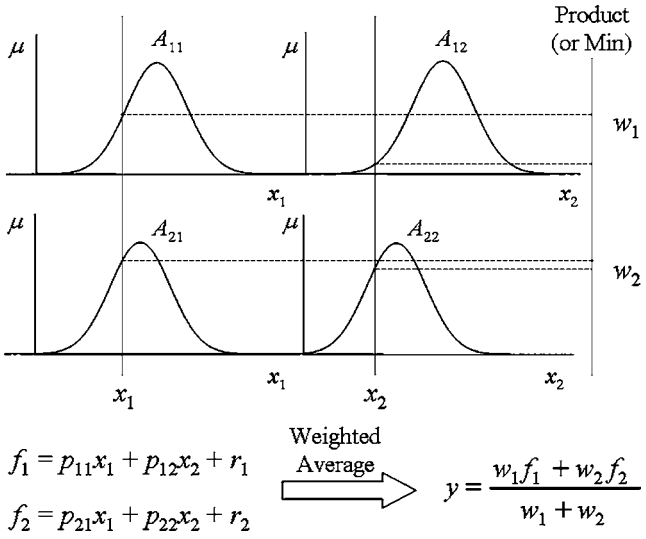


Fig. 2 Two-crisp-input two-rule Sugeno FLS

pressor and turbine) and the corresponding variations of the measured quantities with respect to the values in *new and clean* conditions.

As shown in [12] for NN models, NFSs used for gas turbine diagnostics represent the part of the whole diagnostic system, which is aimed at the evaluation of the relative deviations of the Health Indices with respect to their reference values as a function of the relative deviations of measurements with respect to their values expected in new and clean conditions.

The analyses carried out are aimed at the selection of the most appropriate NFS structure for gas turbine diagnostics, in terms of computational time of the NFS training phase, accuracy, and robustness towards measurement uncertainty during simulations. In particular, *adaptive neuro-fuzzy inference system* (ANFIS) architectures were considered and tested, and their performance was compared to that obtainable by using the NN models. An analysis was also performed in order to identify the most significant ANFIS inputs.

Neuro-Fuzzy Systems

Fuzzy Logic System. A fuzzy logic system (FLS), or fuzzy inference system, is a computing framework based on fuzzy set theory, fuzzy *if-then* rules, and fuzzy reasoning [1].

A *fuzzy set* is a set without a crisp boundary, which extends the definition of a classical set, whose degree of membership can be only 0 or 1, to degrees of membership that can assume values between 0 and 1. The *membership functions* (MFs) are functions mapping each point of the input space to a membership grade between 0 and 1. The transformation process of crisp inputs to fuzzy inputs by means of the MFs is called *fuzzification*.

A fuzzy if-then rule assumes the form

$$\text{if } x \text{ is } A, \text{ then } y \text{ is } B$$

where x and y are linguistic values (such as, for example, "compressor outlet temperature" for x and "compressor efficiency" for y) and A and B are linguistic values (such as, for example, high, low, negligible, etc.). Often "x is A" is called the *antecedent* or *premise* and "y is B" is called the *consequence* or *conclusion*.

Fuzzy reasoning is an inference procedure that derives conclusions from a set of fuzzy if-then rules. The output of fuzzy reasoning is usually a fuzzy set, and, if a crisp output is required, it is necessary to extract a crisp value that best represents the fuzzy output. This process is called *defuzzification*.

Two of the most widespread fuzzy inference systems are the Mamdani and Assilian [14] and Sugeno (also known as Takagi-Sugeno-Kang) [15,16] fuzzy logic systems. In Figs. 1 and 2, an example of the procedures to derive the overall output y from two crisp inputs x_1 and x_2 by using a two-rule Mamdani or Sugeno FLS respectively, is sketched.

In particular, with regard to the Sugeno FLS in Fig. 2, a typical fuzzy rule has the following form:

$$\text{if } x_1 \text{ is } A_1 \text{ and } x_2 \text{ is } A_2, \text{ then } y = f(x_1, x_2)$$

The function $f(x_1, x_2)$ is usually a polynomial in the input variables x_1 and x_2. If f is a first order polynomial, the resulting fuzzy inference system is called *first-order Sugeno FLS*, while, if f is a constant, it is called *zero-order Sugeno FLS* [1]. Both the FLSs considered use the same antecedents, as shown in Figs. 1 and 2. Therefore, the same methods of input space partitioning to form the antecedent of fuzzy rules are applicable to both types of FLSs considered.

A widely used method for input space partitioning is the *grid partition* method [1], though it can be applied only if the number of MFs for each input is small, since it encounters problems for a moderately large number of inputs. In fact, a model with n inputs and n_{MF} MFs for each input results in a number of fuzzy rules $n_{fr}=(n_{MF})^n$. Therefore, as the problem complexity increases, it becomes very difficult to determine the correct set of MFs and fuzzy rules to solve the problem, since a significant amount of time must be spent to correctly tune MFs and adjust fuzzy rules to obtain a good solution; in fact, the capability to relate the rules typically diminishes when approximately $n_{fr} > 15$. For many complex problems, it is impossible to find a proper set of MFs and fuzzy rules to obtain a satisfactory solution [2].

Adaptive Neuro-Fuzzy Inference System (ANFIS). In order to overcome the limits of FLSs in addressing complex problems, NN techniques can be used to learn the behavior of a system from system input-output data. The learned knowledge can then be used to automatically generate the FLS, reducing development time significantly [1,2]. This is possible due to the equivalence between FLSs and feedforward NNs [17]. In this paper, in particular, a class of adaptive NNs functionally equivalent to FLSs, called adaptive neuro-fuzzy inference systems (ANFISs) [1], was considered. A scheme of an ANFIS architecture equivalent to a first-order two-input ($n=2$) single-output Sugeno FLS with two MFs for each input ($n_{MF}=2$) and two rules ($n_{fr}=2$) is reported in Fig. 3.

Five different layers can be recognized [1].

Layer 1 (fuzzification layer). In this layer, x_1 and x_2 crisp inputs are transformed to fuzzy inputs by means of MFs. The outputs of this layer are the antecedents of each rule, represented by membership degrees between 0 and 1. In the present work, Gaussian MFs were considered

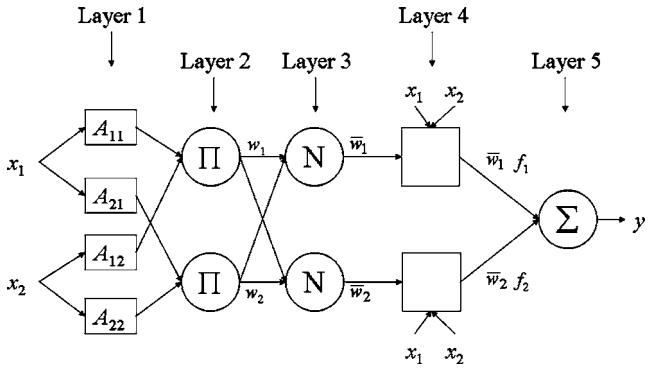


Fig. 3 ANFIS architecture equivalent to a first-order two-input single-output Sugeno FLS with two MFs for each input and two rules

$$\mu_{A_{ij}}(x_j, a_{ij}, \sigma_{ij}) = e^{[(x_j - a_{ij})^2/2\sigma_{ij}^2]}, \quad i = 1, \ldots, n_{\mathrm{MF}}, \quad j = 1, \ldots, n \tag{1}$$

Every node of this layer is an adaptive node defined by the parameters a_{ij} and σ_{ij}, called antecedent parameters.

Layer 2 (rule antecedent layer). Every node of this layer is a fixed node, whose output is the product of the inputs and represents the *firing strength* (weight) of a rule. These outputs depend on the values assumed by the antecedent parameters of layer 1.

Layer 3 (rule strength normalization layer). Every node of this layer is a fixed node whose output is a *normalized firing strength* (normalized weight)

$$(\mathrm{O}_3)_i = \overline{w}_i = \frac{w_i}{\sum_i w_i}, \quad i = 1, \ldots, n_{\mathrm{MF}} \tag{2}$$

Layer 4 (rule consequent layer). Every node of this layer is an adaptive node whose output is

$$(\mathrm{O}_4)_i = \overline{w}_i f_i = \overline{w}_i \left(r_i + \sum_{j=1}^{n} p_{ij} x_j \right), \quad i = 1, \ldots, n_{\mathrm{MF}} \tag{3}$$

Every node is defined by the parameters p_{ij} and r_i, called *consequent parameters*.

Layer 5 (rule inference layer). The single node of this layer is a fixed node that computes the overall output as

$$\mathrm{O}_5 = y = \sum_i \overline{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{4}$$

By considering the ANFIS architecture in Fig. 3, it is evident that it consists of two distinct adaptive parts: the antecedent and the consequent parameters. These two parts can be adapted by different optimization methods. For this work a hybrid learning algorithm was used, since it converges much faster than the original pure back-propagation learning algorithm [1]. For each training set of input and output data:

- in the forward pass from inputs to the output, the algorithm calculates the consequent parameters p_{ij} and r_i by a least-squares method and by considering the antecedent parameters as constant.
- in the backward pass, in which errors propagate backward from the output to inputs, the algorithm updates the antecedent parameters a_{ij} and σ_{ij} of layer 1 by gradient descent and by considering the consequent parameters as constant.

With regard to the problem of input space partitioning to form the antecedent of fuzzy rules, so as to overcome the limits of *grid partition* method for complex problems (i.e., problems with a large number of inputs) highlighted in the previous paragraph, the *subtractive clustering* approach proposed by Chiu [18] was adopted. This method allows the estimation (i) of the cluster centers of input data (and, thus, the determination of the number of clusters), (ii) of the number of MFs of the fuzzification layer and, finally, (iii) of the initial values of antecedent parameters defining the MFs, which are subsequently optimized in the backward passes of the hybrid learning algorithm [1,19]. This represents a strong point with respect to neural networks, for which the number of neurons in the hidden layer, which depends on the number of cluster centers, and the initial values of the connection weights are estimated by means of a trial-and-error procedure or need the use of global optimization methods.

For n_{patt} input data points $(\boldsymbol{x})_i (i = 1, \ldots, n_{\mathrm{patt}})$ in a n-dimensional space, the potential of each $(\boldsymbol{x})_i$ is defined as

$$P_i = \sum_{j=1}^{n_{\mathrm{patt}}} e^{-(4/r_a^2)\|(\boldsymbol{x})_i - (\boldsymbol{x})_j\|^2}, \quad i = 1, \ldots, n_{\mathrm{patt}} \tag{5}$$

where r_a is a positive constant which represents the radius defining a neighborhood and $\| \ \|$ denotes the Euclidean distance. Once the potential of every point has been computed, the subtractive clustering algorithm selects the point with the highest potential as the first cluster center, and then revises the potential of each point according to

$$P_i^{(1)} = P_i^{(0)} - P_1^* e^{-(4/r_b^2)\|(\boldsymbol{x})_i - (\boldsymbol{x})_1^*\|^2}, \quad i = 1, \ldots, n_{\mathrm{patt}} \tag{6}$$

In Eq. (6), $(\boldsymbol{x})_1^*$ and P_1^* are the location of the first cluster center and its potential respectively, while r_b is a positive constant that represents the radius defining the neighborhood, which will have measurable reductions in potential. To avoid obtaining closely spaced cluster centers, r_b is usually set to be somewhat greater than r_a (for example: $r_b = 1.5\, r_a$ in [18] or $r_b = 1.25\, r_a$ in [19]).

When the potentials of all data points have been revised by using Eq. (6), a new point with the highest remaining potential is selected as the second cluster center, and the potential of each point is reduced again by using Eq. (6), in which $(\boldsymbol{x})_1^*$ and P_1^* are substituted with the location of the new cluster center and its potential, respectively. The process of acquiring new cluster centers and revising potential repeats until a criterion for ending the process is verified. This criterion is based on the values assumed by two parameters ε_a and ε_r, which specify the thresholds for accepting or rejecting cluster centers, respectively [18,19].[1]

The number of cluster centers, which corresponds both to the number n_{MF} of the membership functions for each of the n inputs and to the number n_{fr} of fuzzy rules, depends on the values assumed by ε_a, ε_r and r_a. In particular, a large r_a generally results in fewer clusters and, thus, in a coarser model, while small r_a can produce an excessive number of clusters and a model that does not generalize well [19].

When Gaussian MFs (Eq. (1)) are used, the initial values of

[1]The criterion for ending the process is:

If $P_k^* > \varepsilon_a P_1^*$ **then** $(\boldsymbol{x})_k^*$ is accepted as a cluster center and the clustering process is continued
else if $P_k^* < \varepsilon_r P_1^*$ **then** $(\boldsymbol{x})_k^*$ is rejected and the clustering process is stopped
else if $\varepsilon_a P_1^* \geq P_k^* \geq \varepsilon_r P_1^*$ **then**
  if $(d_{\min}/r_a) + (P_k^*/P_1^*) \geq 1$ **then** $(\boldsymbol{x})_k^*$ is accepted as a cluster center and the clustering process is continued
  else $(\boldsymbol{x})_k^*$ is rejected, P_k^* is set to 0, the data with the next highest potential is selected as the new $(\boldsymbol{x})_k^*$ and the test is repeated
  end if
end if

The parameter $d_{\min}$ is the shortest distance between $(\boldsymbol{x})_k^*$ and all previously found cluster centers.

Table 1 Fiat Avio 701F gas turbine main features (ISO conditions, natural gas fuel, no intake and discharge pressure drops)

Power output	255 MW
Overall efficiency	36.0%
Exhaust mass flow rate	574 kg/s
Fuel mass flow rate	13.2 kg/s
Exhaust gas temperature	889 K
Pressure ratio	13.9
Compressor discharge temperature	655 K

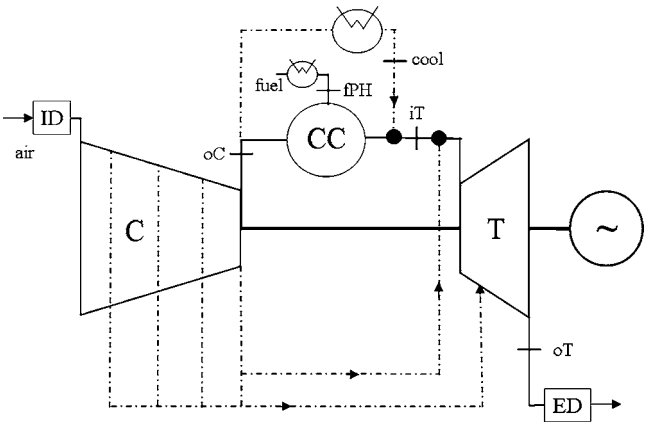


Fig. 4 FIAT Avio 701F gas turbine model

antecedent parameters (which are subsequently optimized in the backward passes of the hybrid learning algorithm), obtained by the subtractive clustering algorithm, are [19]

$$a_{ij} = (x_j^*)_i^*, \quad \sigma_{ij} = \frac{r_a}{\sqrt{8}}, \quad i = 1, \ldots, n_{\mathrm{MF}}, \quad j = 1, \ldots, n$$

where $(x_j^*)_i^*$ is the jth element of ith cluster center $(\boldsymbol{x}^*)_i^*$.

Data for ANFIS Training and Testing

In the development and setup of ANFISs, the same data previously used for the setup of NN models [12] were used. In this manner, it is possible to compare in a proper manner the performance of the ANFISs to that obtainable by using the NN models.

As explained in detail in [12], these data were obtained by using a cycle program, previously calibrated on a FIAT Avio 701F single-shaft gas turbine with variable inlet guide vanes working in the ENEL combined cycle power plant of La Spezia (Italy) [13]. Only data referring to steady-state conditions were generated, since the considered machine is a generator-drive gas turbine, and it usually works in steady-state conditions. The main features and the sketch of the considered gas turbine are reported in Table 1 and in Fig. 4, respectively.

The database, which contains the patterns needed for ANFIS training and testing, was generated by running the cycle program with random values of the inputs, each value being within a prespecified range of variation. The generated data were normalized to be comparable to each other before they were presented to the ANFIS. The normalization was performed with respect to the maximum and minimum value of the data set generated for each variable, i.e.,

$$(Q_i)_{\mathrm{norm}} = \frac{Q_i - (Q_i)_{\min}}{(Q_i)_{\max} - (Q_i)_{\min}}, \quad (Q_i)_{\mathrm{norm}} \in [0,1] \tag{7}$$

In order to simulate field data and, thus, take into account the presence of measurement errors, the generated data were then corrupted with random errors, based on measurement uncertainties which are typical for a standard instrumentation category [12,20].

In Table 2, the inputs with the range of variation and the outputs of the cycle program, and the inputs with the considered measurement uncertainty values and the outputs of the ANFISs are highlighted. In particular, the diagnostic system consists of a number of ANFISs equal to the number of Health Indices used for gas turbine diagnostics, each ANFIS calculating one Health Index as a function of the inputs. In the present case, the considered Health Indices are the ones related to compressor and turbine efficiencies and corrected mass flow rates (η_{C}, η_{T}, μ_{C}, and μ_{T}). These param-

Table 2 Inputs (with their range of variation) and outputs of the cycle program (CP) and inputs (with the measurement uncertainty values) and outputs of the ANFISs

	Quality	Unit	Variation range	CP I/O	ANFIS I/O	Measurement uncertainty
p_{amb}	Ambient pressure	kPa	[98; 103]	I	I	±0.63 kPa
T_{amb}	Ambient temperature	K	[278.15; 308.15]	I	I	±0.3 K
RH	Relative humidity	%	[20; 100]	I	I	±2.0%
Δp_{ID}	Pressure drop at the compressor inlet	kPa	[0; 2.5]	I	I	±0.10 kPa
Δp_{ED}	Pressure drop at the turbine outlet	kPa	[0; 4.0]	I	I	±0.10 kPa
N	Rotational speed	rpm	[2990; 3010]	I	I	±30 rpm
T_{cool}	Cooling flow temp. after the exchanger	K	[453.15; 493.15]	I	I	±1.0 K
T_{fPH}	Temperature of the fuel after heating	K	[288.15; 383.15]	I	I	±1.0 K
LHV	Fuel lower heating value	kJ/kg	[40,000; 50,000]	I	I	±3.00%[a]
A_{IGV}	IGV angular position	deg	[−4;25]	I	I	±1.0 deg
T_{iT}	Turbine inlet temperature	K	[1150; 1750]	I	-	
η_{C}	Compressor efficiency health index	-	[0.95; 1.05]	I	O	
μ_{C}	Comp. corrected mass flow health index	-	[0.90; 1.10]	I	O	
η_{T}	Turbine efficiency health index	-	[0.95; 1.05]	I	O	
μ_{T}	Turbine corrected mass flow health index	-	[0.90; 1.10]	I	O	
$P\mathrm{GT}$	Gas turbine power	kW		O	I	±2.00%[a]
p_{oC}^*	Nondimensional compressor outlet pressure	-		O	I	±1.43%[a]
T_{oC}^*	Nondimensional compressor outlet temperature	-		O	I	±0.23%[a]
M_{f}^*	Nondimensional fuel mass flow rate	-		O	I	±1.50%[a]
T_{oT}^*	Nondimensional turbine outlet temperature	-		O	I	±0.90%[a]

[a]Percentage of the reading.

eters do not depend on the gas turbine operating point but are only sensitive to the gas turbine health state [7,12,21–23]. According to [24], this combination of characteristic parameters represents the best solution for the number and kind of available measurements for the considered machine.

The inputs of each ANFIS are the parameters that unequivocally determine the operating point at which the gas turbine is working and four further parameters used for gas turbine diagnostics, which are the ratios, with respect to the values expected in new and clean conditions, of the measured values of compressor outlet pressure p^*_{oC} and temperature T^*_{oC}, fuel mass flow rate M^*_f, and turbine outlet temperature T^*_{oT}.

Influence Analysis of ANFIS Configuration Parameters

The ANFISs that were developed in this work are functionally equivalent to a first-order, 15-input single-output Sugeno FLS with Gaussian membership functions. Because of the high number of ANFIS inputs, the *subtractive clustering* approach was used [18,19], since, as explained in the previous paragraph, the *grid partition* method is not suitable for problems with a large number of inputs. An analysis was nevertheless conducted to evaluate the capability of the *grid partition* approach for ANFISs with a reduced number of inputs and results are reported at the end of this section.

In order to identify the most appropriate ANFIS structure for gas turbine diagnostics, in terms of computational time of the ANFIS training phase, accuracy and robustness against measurement uncertainty, a sensitivity analysis was carried out with respect to the main ANFIS configuration parameters. The ANFIS parameters whose influence was analyzed in the paper are

- clustering algorithm parameters.
- optimal training set dimension, i.e., optimal number of patterns that have to be used for the ANFIS training process.
- number of ANFIS inputs, i.e., choice of the most significant inputs.
- comparison between *subtractive clustering* and *grid partition* approaches for ANFISs with a reduced number of inputs.

Furthermore, the ANFIS robustness against measurement uncertainty was evaluated for all the considered cases.

The parameter used for the comparison of the ANFISs is the root-mean-square error RMSE made by each ANFIS on the whole set of test data in the calculation of the respective output $(\eta_C, \mu_C, \eta_T, \mu_T)$:

$$\left(\mathrm{RMSE} = \sqrt{\frac{1}{n_{\mathrm{patt}}}\sum_{i=1}^{n_{\mathrm{patt}}}\left(\frac{t_i - y_i}{t_i}\right)^2}\right)_j, \quad j = 1, \ldots, n_{\mathrm{HI}} \tag{8}$$

where t_i are the target outputs, y_i the computed outputs, n_{patt} is the number of patterns used for ANFIS testing, and n_{HI} is the number of the Health Indices (and, so, the number of single-output ANFISs) considered for gas turbine diagnostics.

In addition to the root-mean-square error RMSE, ANFISs were also compared by means of a parameter called success rate (SR). It has been considered that a success in the simulation of a computed output y_j is achieved when it results that

$$|e|_j = |t - y|_j \leq \delta, \quad j = 1, \ldots, n_{\mathrm{HI}} \tag{9}$$

where δ is a chosen threshold. This allows the calculation of the success rate for each output, defined as the percentage ratio between the number of successes and the number of test patterns n_{patt}.

Finally, the developed ANFISs were compared to each other with respect to training phase computational time. All computational times reported in the paper refer to calculations performed by using a PC with a Pentium 3 800 MHz CPU.

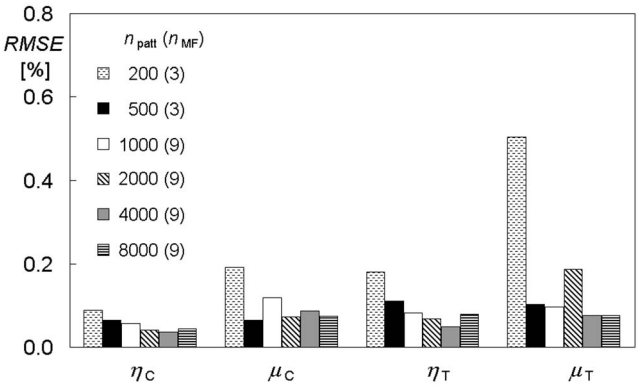


Fig. 5 Influence of the number of training patterns

Clustering Algorithm Parameters. A preliminary analysis was conducted in order to identify the optimal values of the clustering algorithm parameters r_a and r_b. Even if, as mentioned previously, r_b is usually set to be somewhat greater than r_a [18,19], the results of the analysis showed that, for the particular problem considered, a good solution can be obtained by assuming $r_a = r_b = 1$. Thus, this couple of values was adopted throughout the rest of the paper.

An analysis was also conducted to evaluate the influence of the number n_{MF} of MFs for each of the 15 inputs, which, as explained in the previous paragraph, depends on the values imposed on the clustering algorithm parameters ε_a and ε_r [18,19]. ANFISs with a number of MFs for each of the 15 inputs varying from 3 to 9 were taken into consideration ($n_{\mathrm{MF}} = 3, 5, 7, 9$). It was found that, when the number of training patterns is lower than 1000, the optimal number of MFs for each of the 15 inputs is 3, whereas, when the number of training patterns is equal to or higher than 1000, the optimal n_{MF} is 9. Thus, the analyses carried out in the next paragraph were performed by using $n_{\mathrm{MF}} = 3$ for ANFISs trained with less than 1000 training patterns, while $n_{\mathrm{MF}} = 9$ was adopted for ANFISs trained with at least 1000 training patterns.

Number of Training Patterns. The influence of the number of patterns used to train the ANFISs on the simulation error was evaluated by comparing the response of different ANFISs, which were trained by using different training sets with a number of patterns from 200 to 8000. The RMSE of each ANFIS was evaluated by running the ANFIS in order to reproduce 40,000 patterns (not including the patterns used for training). The analysis of the results, presented in Fig. 5, shows that, even when a low number of training patterns is used (i.e., $n_{\mathrm{patt}} = 500$), RMSE is very small ($\sim 0.1\%$). Thus, it is not advantageous to use a higher number of training patterns due to the increase in the training phase computational time. In fact, the overall time required for the training of the four single-output ANFISs is 10 min for $n_{\mathrm{patt}} = 500$, while it becomes 3.5 h for $n_{\mathrm{patt}} = 4000$.

The comparison to the results obtained in [12] by using multi-input multioutput (MIMO) NN structures (i.e., NNs calculating all the Health Indices) and multi-input single-output (MISO) NNs (each of them calculating a single Health Index) is reported in Fig. 6. It can be noted that, for all outputs, the RMSE for ANFISs is always lower than the corresponding RMSE for NNs, for any given number of NN outputs and of training patterns. These results are confirmed by calculating the success rate for the three models. An example, in the case of a number of training patterns equal to 1000 and a threshold δ equal to 0.5%, is reported in Table 3. In particular, it can be observed that for ANFIS the success rate is close to 100% for all the four outputs. The advantage of ANFISs with respect to NNs is evident, mainly for the determination of both corrected mass flow rate Health Indices μ_C and μ_T. The

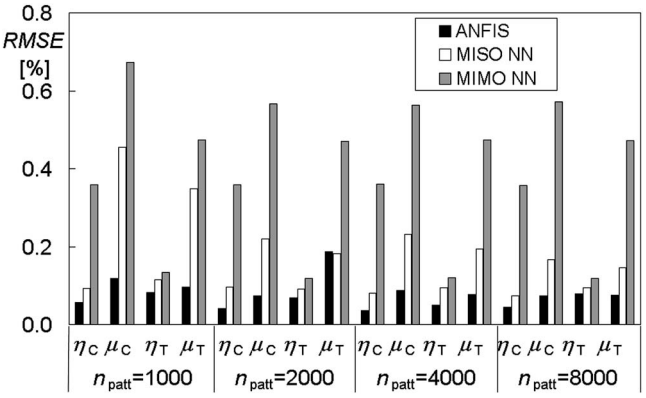


Fig. 6 Influence of the number of training patterns: comparison of ANFISs (n_{MF}=9), MISO, and MIMO neural networks

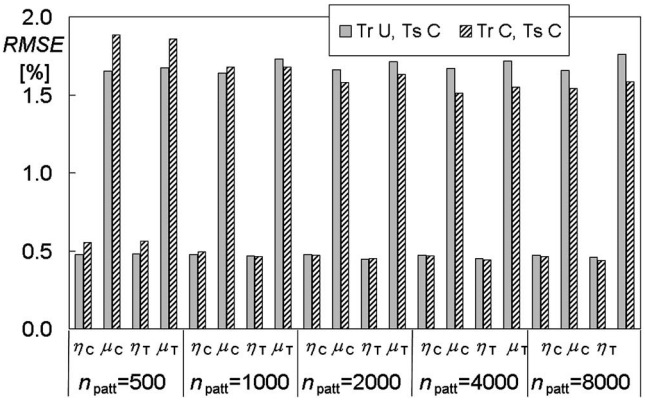


Fig. 7 Influence of the number of training patterns in the presence of measurement errors: comparison between ANFISs trained with data not corrupted with measurement errors (Tr U) and with data corrupted with measurement errors (Tr C) (n_{MF} =3 in all cases)

comparison of the overall computational time required for the training phase is also in favor of ANFISs. For example, in the case of a number of training patterns equal to 1000, the time required for training is equal to about 45 min, 2.5 h, and 8 h for ANFIS, MIMO NN, and MISO NN, respectively.

Measurement Uncertainty. Different ANFISs were considered to assess their robustness towards measurement uncertainty. The analyses were focused on the evaluation of the influence of

- the number of patterns used for the training phase; in particular, 500, 1000, 2000, 4000 and 8000 patterns were used for training.
- the presence of measurement errors; two situations were investigated:
  - presence of measurement errors only in the test patterns, i.e., training patterns are not corrupted with measurement errors, while test patterns are corrupted with measurement errors (case Tr U, Ts C).
  - presence of measurement errors both in training and test patterns, i.e., both training and test patterns are corrupted with measurement errors (case Tr C, Ts C).

In the test phase, the RMSE of each ANFIS was evaluated by running the ANFIS in order to reproduce the 40,000 test patterns, which, as usual, do not include the patterns used for training. For test data affected by measurements errors, the optimal number of MFs for each of the 15 inputs proved to be three, independent of the number of training patterns and of the presence of measurement errors in the training patterns. In fact, for data corrupted by measurement errors, the relationship between inputs and outputs is weaker ("less deterministic") than for data not corrupted by measurement errors, so that the optimal number of cluster centers is low, and thus, the model configuration is leaner. This same result can be also extended to NNs, for which the number of neurons in the hidden layer plays the role of the number of cluster centers. Therefore, ANFISs with n_{MF}=3 were always adopted for the analyses performed in this section. The results are reported in Fig. 7, where it can be noted that:

- RMSE values on efficiencies η_C and η_T are lower than RMSE on corrected mass flow rates μ_C and μ_T. In fact, RMSE values of ~0.5% for η_C and η_T are reached by using only 500 training patterns, whereas RMSE values for μ_C and μ_T remain in the range 1.5–1.9% even if 8000 training patterns are used.
- RMSE decreases as the number of training patterns increases, though its influence is appreciable only when passing from 500 to 1000 training patterns. In fact, when over 1000 training patterns are used, such a benefit becomes negligible.
- RMSE values for ANFISs trained by using data corrupted with measurement errors (Tr C) are lower than those obtained by using data not corrupted with measurement errors (Tr U) for all outputs, if at least 2000 training patterns are considered.

The comparison to the results of a similar analysis conducted in [12] is reported in Fig. 8, where the results refer to ANFISs (with n_{MF}=3), MIMO, and MISO neural networks all trained with data corrupted with measurement errors.

It can be observed that, for all outputs, the RMSE for ANFISs is

Table 3 Success rate for ANFISs, MISO, and MIMO neural networks (number of training patterns=1000; δ=0.5%) (data not corrupted by measurement errors)

	Success rate SR (%)			
	η_C	μ_C	η_T	μ_T
ANFIS	99.95	99.27	99.96	99.73
MISO NN	99.89	82.69	99.79	89.80
MIMO NN	91.59	37.32	99.76	70.07

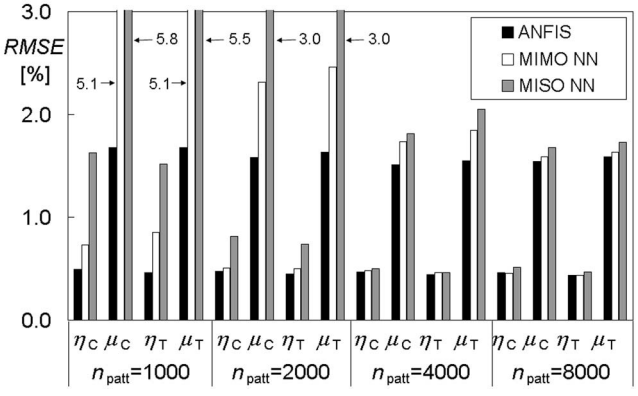


Fig. 8 Influence of the number of training patterns in the presence of measurement errors: comparison of ANFISs (n_{MF}=3), MIMO, and MISO NNs, all trained with data corrupted with measurement errors

Table 4 Difference between the RMSE made by the ANFISs fed with 14 inputs ($RMSE_{14}$) and the RMSE made by the original ANFISs fed with 15 inputs ($RMSE_{15}$)

"Excluded" Input	$\Delta RMSE = RMSE_{14} - RMSE_{15}$			
	η_C	μ_C	η_T	μ_T
p_{amb}	~	~	~	~
T_{amb}	↑	~	~	~
RH	~	~	~	~
Δp_{ID}	~	~	~	~
Δp_{ED}	~	~	~	~
N	~	~	~	~
T_{cool}	~	~	~	~
T_{fPH}	~	~	~	~
LHV	~	~	~	~
A_{IGV}	↑	↑	~	~
P_{GT}	↑↑	↑↑	↑↑	↑
p^*_{oC}	↑↑↑	↑↑	↑↑	↑↑↑
T^*_{oC}	↑↑↑	↑↑	↑↑	↑↑
M^*_f	~	↑↑↑	↑↑	↑↑↑
T^*_{oT}	↑	↑↑↑	~	↑↑↑
~	$RMSE_{14} - RMSE_{15} < 10\%$			
↑	$10\% < RMSE_{14} - RMSE_{15} < 30\%$			
↑↑	$30\% < RMSE_{14} - RMSE_{15} < 50\%$			
↑↑↑	$RMSE_{14} - RMSE_{15} > 50\%$			

always lower than the corresponding RMSE for NNs, for any given number of training patterns. Since RMSE values for NNs decrease by increasing the number of training patterns more rapidly than RMSE values for ANFISs, the above-described benefit reduces when a high number of training patterns is used.

The advantage of ANFISs with respect to NNs is also verified by comparing the overall computational time required for training. In fact, the time required for ANFIS training is equal to about 20 min or 50 min if 2000 or 4000 training patterns are used, whereas for the MIMO or the MISO NNs it would be 3.5 h and 8 h or 13 h and 17.5 h, respectively. It should be noted that the time required for training the ANFIS developed in this section is low, if compared to the ANFIS developed in the previous section, since the considered ANFIS has only three MFs for each of the 15 inputs.

Number of Inputs. The influence of the number of ANFIS inputs on the simulation error was evaluated by comparing the response of different ANFISs, which were set up by considering as input variables only 5 or 6 out of the 15 previously considered input variables. The RMSE of each ANFIS was evaluated by running the ANFIS to reproduce 40,000 patterns corrupted with measurement errors, not including the patterns used for training, as done before.

The first step of the analysis was the assessment of the significance of each input. This was performed according to the following procedure:

- selection of the four 15-input ANFISs trained on data not corrupted with measurement errors characterized by: $r_a = r_b = 1$, $n_{MF} = 3$, 4000 training patterns.
- execution of 15 different simulations on the 40,000 test patterns, by feeding the ANFISs with only 14 inputs and by excluding in turn a different input for each of the 15 simulations; to exclude each input the corresponding coefficients of the consequent parts of fuzzy rules was set at zero.
- calculation of ΔRMSE as the difference between the RMSE made by the ANFISs fed with 14 inputs ($RMSE_{14}$) and the RMSE made by the original ANFIS fed with 15 inputs ($RMSE_{15}$).

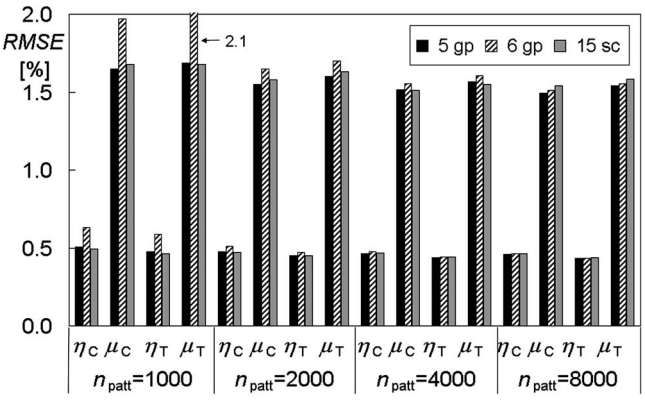


Fig. 9 Influence of the number of inputs (5, 6, and 15) and of the input space partitioning algorithm (gp: grid partition; sc: subtractive clustering); training data corrupted with measurement errors

The results are presented synoptically in Table 4, for any given "excluded" input and for any output. The high ΔRMSE values (up to 50%) are due to the fact that the original ANFIS architectures are characterized by 15 inputs, while, in these simulations, the ANFISs were fed with only 14 inputs. In fact, it was verified that, when new ANFISs with a 14-input architecture are instead set up and trained, the RMSE values on test patterns are of the same order of magnitude as the ones obtained by using the 15-input ANFISs fed with 15 inputs.

It can be observed that, for all outputs, in accordance with thermodynamic considerations, the most significant inputs (i.e., the ones that, if "excluded" from ANFIS, would lead to the largest errors) are (i) gas turbine power P_{GT} and (ii) nondimensional values of the measurable quantities used for gas turbine diagnostics (compressor outlet pressure p^*_{oC} and temperature T^*_{oC}, fuel mass flow rate M^*_f, and turbine outlet temperature T^*_{oT}). Another significant input is the IGV angular position A_{IGV}, especially for the determination of compressor Health Indices. Thus, the following ANFISs, trained with 1000, 2000, 4000, and 8000 training patterns corrupted with measurement errors, were developed:

- five-input ANFISs (P_{GT}, p^*_{oC}, T^*_{oC}, M^*_f, and T^*_{oT}), generated by using the *grid partition* algorithm.
- six-input ANFISs (P_{GT}, p^*_{oC}, T^*_{oC}, M^*_f, and T^*_{oT}, and A_{IGV}), generated by using the grid partition algorithm.
- six-input ANFISs (P_{GT}, p^*_{oC}, T^*_{oC}, M^*_f, T^*_{oT}, and A_{IGV}), generated by using the *subtractive clustering* algorithm.

Then, the above-mentioned ANFISs were compared to each other and with the 15-input ANFISs generated by using the subtractive clustering algorithm. It can be noted that

- when the subtractive clustering algorithm is used, the total number of fuzzy rules is $n_{fr} = n_{MF}$; a number n_{MF} of MFs for each input equal to 3 was used both for 15 and 6-input ANFISs.
- when the grid partition algorithm is used, the total number of fuzzy rules is $n_{fr} = (n_{MF})^n$, where n is the number of inputs. A number n_{MF} of MFs for each input equal to 2 was used both for five and six-input ANFISs.

The results of the analysis of the influence of the number of inputs are presented in Fig. 9, where the ANFISs with a reduced number of inputs (five or six) were generated by using the grid partition algorithm, while the 15-input ANFISs were generated by using the subtractive clustering algorithm. Two considerations can be derived:

- For the ANFISs with a reduced number of inputs, RMSE

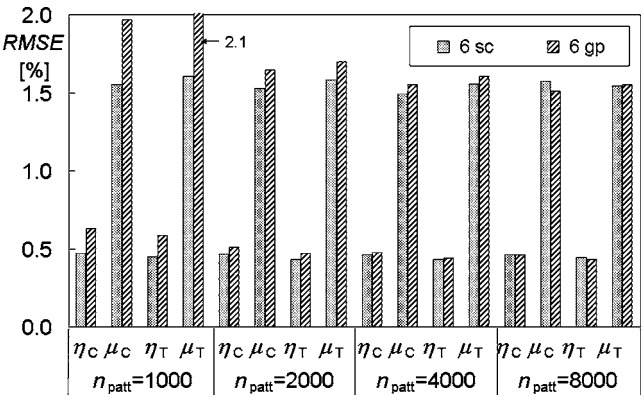


Fig. 10 Influence of the input space partitioning algorithm (gp: grid partition; sc: subtractive clustering) for six-input ANFIS trained with data corrupted with measurement errors

also decreases as the number of training patterns increases, though such an effect is more considerable for the six-input ANFIS.

- When a high number of training patterns is used (i.e., 8000), both five and six-input ANFISs enable lower RMSE to be obtained with respect to the 15-input ANFIS.

The results of the comparison between two six-input ANFISs trained with data corrupted with measurement errors, generated by using the subtractive clustering algorithm (6 sc) and by using the grid partition algorithm (6 gp), respectively, are reported in Fig. 10. It can be observed that when less than 4000 training patterns are used, the subtractive clustering algorithm is preferable with respect to the grid partition method, whereas for a high number of training patterns the two algorithms are comparable.

Finally, the SRs for three different thresholds (δ=0.5%, 1.0%, 2.0%) obtained by using MIMO and MISO NNs, 15-input ANFISs generated by using the subtractive clustering algorithm, six-input ANFISs generated by using both subtractive clustering and grid partition algorithms and five-input ANFISs generated by using the grid partition algorithm, are reported in Table 5. These SRs are relative to models all trained with 4000 training patterns corrupted with measurement errors. The results in Table 5 substantially confirm the ones reported in Figs. 8–10, i.e., as follows:

- The performances (in terms of higher SR) obtained by using the ANFISs are always slightly better than the ones obtainable by means of MIMO and MISO neural networks.
- For 4000 training patterns, the SRs obtained by using 15-input ANFISs generated by using the subtractive clustering algorithm, six-input ANFISs generated by using both subtractive clustering and grid partition algorithms and five-input ANFISs generated by using the grid partition algorithm are comparable to each other.

Discussion

The diagnostic system considered in the paper consists of four 15-input ANFISs, each of them calculating as output one of the four Health Indices (efficiencies and corrected mass flow rates of compressor and turbine) as a function of the inputs. The most interesting results of the developed analyses can be summarized as follows:

- When data not corrupted by measurement errors are used for ANFIS training and testing, the best compromise between computational time of the ANFIS training phase and accuracy during simulations seems to be an ANFIS with three MFs for each of the 15 inputs trained by using at least 500 training patterns.

Table 5 Success rates for: MIMO and MISO NNs, 15, 6, and 5-input ANFISs (gp: grid partition; sc: subtractive clustering), all trained with 4000 training patterns corrupted with measurement errors (SR=100% set in bold)

		Success rate SR (%)		
	Threshold δ (%)	0.5	1.0	2.0
η_C	MISO NN	66	96	**100**
	MIMO NN	68	97	**100**
	ANFIS (15 sc)	69	98	**100**
	ANFIS (6 sc)	69	98	**100**
	ANFIS (6 gp)	68	97	**100**
	ANFIS (5 gp)	69	98	**100**
μ_C	MISO NN	22	42	73
	MIMO NN	23	44	75
	ANFIS (15 sc)	24	47	80
	ANFIS (6 sc)	25	48	81
	ANFIS (6 gp)	24	47	79
	ANFIS (5 gp)	25	48	80
η_T	MISO NN	71	97	**100**
	MIMO NN	71	97	**100**
	ANFIS (15 sc)	72	98	**100**
	ANFIS (6 sc)	74	99	**100**
	ANFIS (6 gp)	72	98	**100**
	ANFIS (5 gp)	73	99	**100**
μ_T	MISO NN	21	40	70
	MIMO NN	22	43	74
	ANFIS (15 sc)	25	47	80
	ANFIS (6 sc)	24	46	79
	ANFIS (6 gp)	24	46	78
	ANFIS (5 gp)	25	47	79

- When data corrupted by measurement errors are used for ANFIS testing, it is advantageous to use ANFISs with three MFs for each of the 15 inputs trained with at least 1000 training patterns affected by measurement errors.

In all the cases analyzed, the performances (in terms of lower RMSE, higher success rate and shorter time spent for the training phase) obtained by using the ANFISs proved to be better than the ones obtainable by MIMO and MISO neural networks trained and tested on the same data.

Finally, the analysis on the influence of the number of ANFIS inputs has shown that, when at least 2000 training patterns are used, both five and six-input ANFISs enable RMSE values close to the ones of 15-input ANFISs to be obtained. Regarding the approach for input space partitioning, it has been observed that when less than 2000 training patterns are used, the subtractive clustering algorithm is usually preferable with respect to the grid partition method, whereas for a high number of training patterns the two algorithms are comparable.

Conclusions

In the paper, adaptive neuro-fuzzy inference systems (ANFISs) for gas turbine diagnostics have been studied and developed. The data used for training and testing the ANFISs were generated by running a cycle program, calibrated on a 255 MW single-shaft gas turbine. The analyses carried out were aimed at the selection of the most appropriate ANFIS structure for gas turbine diagnostics, in terms of computational time of the ANFIS training phase, accuracy, and robustness towards measurement uncertainty during simulations.

As a general conclusion, the results obtained show that ANFISs are robust with respect to measurement uncertainty, and, in all the cases analyzed, the performance (in terms of accuracy during simulations and time spent for the training phase) proved to be

better than that obtainable by MIMO and MISO neural networks trained and tested on the same data. Moreover, when ANFISs are fed with data affected by measurement errors, ANFISs with a simple structure (i.e., characterized by a low number of membership functions) and trained with a high number of patterns corrupted with measurement errors are recommended. Thus, ANFIS models represent a quite easy-to-implement solution to support the diagnostics of gas turbines, and in some cases, they perform better than other artificial intelligence techniques, such as neural networks.

Acknowledgment

The work was carried out with the support of the MiUR (Italian Ministry of University and Scientific & Technological Research).

Nomenclature

a_{ij} = antecedent parameter (Eq. (1))
A_{IGV} = inlet guide vane angular position
ANFIS = adaptive neuro-fuzzy inference system
C = corrupted (with measurement error)
e = $t-y$ error
FLS = fuzzy logic system
gp = grid partition
I = input
LHV = fuel lower heating value
M = mass flow rate
MF = membership function
MIMO = multi-input/multioutput
MISO = multi-input/single-output
n = number of inputs
n_{fr} = number of fuzzy rules
n_{HI} = number of Health Indices considered for gas turbine diagnostics
n_{MF} = number of membership functions for each input
n_{patt} = number of patterns
N = rotational speed
NFS = neuro-fuzzy system
NN = neural network
O = output
p = pressure
p_{ij} = consequent parameter (Eq. (3))
P = power, potential
Q = quantity
r_a, r_b = radii of influence of a clustering center
r_i = consequent parameter (Eq. (3))
RH = relative humidity
RMSE = root-mean-square error
sc = subtractive clustering
SR = success rate
t = expected target output, time
T = temperature
Tr = training data
Ts = testing data
U = uncorrupted (without measurement error)
w_i = rule *firing strength* (weight)
$\overline{w}_i$ = normalized *firing strength* (Eq. (2))
$\boldsymbol{x}$ = $[x_1, \ldots, x_n]^T$ input vector
y = computed output
δ = threshold
Δ = variation
$\varepsilon_{a(r)}$ = thresholds for accepting (rejecting) cluster centers
η = efficiency health index
μ = corrected mass flow rate health index, Gaussian membership function (Eq. (1))
σ_{ij} = antecedent parameter (Eq. (1))

Subscripts and Superscripts

* = nondimensional value, cluster center
amb = ambient
cool = cooling flow after the heat exchanger
C = compressor
ED = exhaust duct
f = fuel
GT = gas turbine
i = inlet section
ID = inlet duct
max = maximum
min = minimum
norm = normalized value
o = outlet section
PH = post heating
T = turbine

References

[1] Jang, J.-S. R., Sun, C.-T., and Mizutani, E., 1997, *Neuro-Fuzzy and Soft Computing*, Prentice-Hall, Englewood Cliffs, NJ.
[2] Jain, L. C., and Martin, N. M., 1998, *Fusion of Neural Networks, Fuzzy Sets, and Genetic Algorithms*, CRC Press, Boca Raton.
[3] Torella, G., and Lombardo, G., 1996, "Neural Networks for the Diagnostics of Gas Turbine Engines," ASME Paper No. 96-TA-39.
[4] Kanelopoulos, K., Stamatis, A., and Mathioudakis, K., 1997, "Incorporating Neural Networks Into Gas Turbine Performance Diagnostics," ASME Paper No. 97-GT-35.
[5] Volponi, A. J., DePold, H. R., Ganguli, R., and Daguang, C., 2000, "The Use of Kalman Filter and Neural Networks Methodologies in Gas Turbine Performance Diagnostics: A Comparative Study," ASME Paper No. 2000-GT-0547.
[6] Romessis, C., Stamatis, A., and Mathioudakis, K., 2001, "A Parametric Investigation of the Diagnostic Ability of Probabilistic Neural Networks on Turbofan Engines," ASME Paper No. 2001-GT-0011.
[7] Bettocchi, R., Spina, P. R., and Torella, G., 2002, "Gas Turbine Health Indices Determination by Using Neural Networks," ASME Paper No. GT-2002-30276.
[8] Ganguli, R., 2003, "Application of Fuzzy Logic for Fault Isolation of Jet Engines," ASME J. Eng. Gas Turbines Power, **125**, pp. 617–623.
[9] Arriagada, J., Genrup, M., Loberg, A., and Assadi, M., 2003, "Fault Diagnosis System for an Industrial Gas Turbine by Means of Neural Networks," *Proc. of International Gas Turbine Congress 2003 (IGTC'03)*, Tokyo, Nov. 2–7, GTSJ, Tokyo, Paper No. IGTC2003Tokyo TS-001.
[10] Sampath, S., and Singh, R., 2004, "An Integrated Fault Diagnostics Model Using Genetic Algorithm and Neural Networks," ASME Paper No. GT-2004-53914.
[11] Ganguli, R., Verma, R., and Roy, N., 2004, "Soft Computing Application for Gas Path Fault Isolation," ASME Paper No. GT-2004-53209.
[12] Bettocchi, R., Pinelli, M., Spina, P. R., and Venturini, M., 2007, "Artificial Intelligence for the Diagnostics of Gas Turbines—Part I: Neural Network Approach," ASME J. Eng. Gas Turbines Power, **129**(3), pp. 711–719.
[13] Bettocchi, R., Pínelli, M., Venturini, M., Spina, P. R., Bellagamba, S., and Tirone, G., 2002, "Procedura di Calibrazione del Programma per la Diagnosi Funzionale dei Turbogas della Centrale a Ciclo Combinato di La Spezia," *Proc. 57th Congresso Nazionale ATI*, Pisa, ETS ed., September 17–20, pp. III-B-3–III-B-12 (in Italian).
[14] Mamdani, E. H., and Assilian, S., 1975, "An Experiment in Linguistic Synthesis With a Fuzzy Logic Controller," J. Man-Mach. Stud., **7**, pp. 1–13.
[15] Takagi, T., and Sugeno, M., 1985, "Fuzzy Identification of Systems and Its Applications to Modeling and Control," IEEE Trans. Syst. Man Cybern., **15**, pp. 116–132.
[16] Sugeno, M., and Kang, G. T., 1988, "Structure Identification of Fuzzy Model," J. Fuzzy Sets Syst., **28**, pp. 15–33.
[17] Li, H.-X., and Chen, C. L. P., 2000, "The Equivalence Between Fuzzy Logic Systems and Feedforward Neural Networks," IEEE Trans. Neural Netw., **11**, pp. 356–365.
[18] Chiu, S. L., 1994, "Fuzzy Model Identification Based on Cluster Estimation," J. Intell. Fuzzy Syst., **2**, pp. 267–278.
[19] Chiu, S. L., 1997, "Extracting Fuzzy Rules From Data for Function Approximation and Pattern Classification," *Fuzzy Information Engineering: A Guided Tour of Applications*, Wiley, New York, Chap. 9.
[20] Bettocchi, R., Pinelli, M., Spina, P. R., Venturini, M., and Burgio, M., 2004, "Set Up of a Robust Neural Network for Gas Turbine Simulation," ASME

Paper No. GT2004-53421.
[21] Stamatis, A., Mathioudakis, K., and Papailiou, K. D., 1990, "Adaptive Simulation of Gas Turbine Performance," ASME J. Eng. Gas Turbines Power, **112**, pp. 168–175.
[22] Bettocchi, R., and Spina, P. R., 1999, "Diagnosis of Gas Turbine Operating Conditions by Means of the Inverse Cycle Calculation," ASME Paper No. 99-GT-185.
[23] Pinelli, M., Spina, P. R., and Venturini, M., 2003, "Optimized Operating Point Selection for Gas Turbine Health State Analysis by Using a Multi-Point Technique," ASME Paper No. GT2003-38191.
[24] Bettocchi, R., Pinelli, M., Spina, P. R., and Venturini, M., 2003, "Statistical Analyses to Improve Gas Turbine Diagnostics Reliability," *Proc. of 8th International Gas Turbine Congress*, Nov. 2–7, Tokyo, GTSJ, Tokyo, pp. IGTC2003Tokyo TS-004.

Manuj Dhingra
Yedidia Neumeier
J. V. R. Prasad

School of Aerospace Engineering,
Georgia Institute of Technology,
Atlanta, GA 30332

Andrew Breeze-Stringfellow
Hyoun-Woo Shin
Peter N. Szucs

GE Aviation,
Cincinnati, OH 45215

A Stochastic Model for a Compressor Stability Measure

A stability measure rooted in the unsteady characteristics of the flow field over the compressor rotor has been previously developed. The present work explores the relationship between the stochastic properties of this measure, called the correlation measure, and the compressor stability boundary. A stochastic model has been developed to gauge the impact of the correlation measure's stochastic nature on its applicability to compressor stability management. The genesis of this model is in the fundamental properties of a specific stochastic process, one that is created by the threshold crossings of a random process. The model validation utilizes data obtained on three different axial compressor facilities. These include a single-stage low-speed axial compressor, a four-stage low-speed research compressor, and an advanced technology demonstrator high-speed compressor. This paper presents details of the model development and validation, as well as closed loop experimental results to demonstrate correlation measure's usefulness in compressor stability management. [DOI: 10.1115/1.2718231]

This paper has received the Best Paper Award of the Controls, Diagnostics & Instrumentation Committee, Turbo Expo '06

Introduction

Compressor aerodynamic instabilities, rotating stall and surge, impose limits on the safe operation of a gas turbine engine. Compressor performance is usually represented via a map of pressure ratio versus mass flow rate iso-speed lines. On this map (Fig. 1), the onset of compressor instabilities is denoted by the surge line, while the operating line is the locus of the steady state operating points. The term "stall margin" is used to denote the distance to the surge line. The stall margin requirement at the design operating point tends to be conservative, in part due to the effects such as aging, speed transients, thermal transients, engine control tolerances, and inlet distortions. Along with a conservative stall margin, a fuel acceleration schedule is employed, which keeps the operating line excursions during engine transients well below the estimated surge line.

An active stability management system can alleviate the impact of surge line uncertainties and improve gas turbine operation in different ways. For some designs it may happen that the point of maximum efficiency lies above the operating line. Even when this is not the case, current designs incur excess solidity to achieve stall margin requirements, and sacrifice some efficiency in the process. The reduction in stall margin requirement can enable improvements in efficiency and a reduction in weight. With higher loading, it may even be possible to reduce the number of stages and thus reduce weight as well as engine part count. Either would lead to benefits from the commercial operations perspective.

Alternatively, active stability management can improve engine dynamic response by utilizing all of the available stall margin. Certain maneuvers, such as aborted carrier landings by fixed wing aircrafts, or autorotation recovery in helicopters, demand a sharp increase in propulsion system output. The open loop scheduler, while keeping the engine safe even in the worst-case scenario, limits the engine response time. Thus, the successful application of active compressor control could also be beneficial in military applications.

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received July 5, 2006; final manuscript received November 17, 2006. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2006: Land, Sea and Air (GT2006), May 8–11, 2006, Barcelona, Spain.

A model for the poststall transients first appeared about two decades ago [1,2]. On the basis of this model, Epstein et al. [3] evaluated the benefits of active control when applied to compressor aerodynamic instabilities. They postulated that only low authority control was sufficient to suppress stall at its inception. The idea of active compressor control spurred numerous efforts aimed at stabilizing part of the unstable region [4–8]. On the other hand, prompted by the severity of the aerodynamic instabilities, researchers sought to reduce the stall margin requirements, while completely avoiding the regime susceptible to these instabilities. In this context, precursor-based schemes have been proposed. Tryfonidis et al. [9] found prestall modal waves in several compressors and developed the concept of the traveling wave energy as a criteria for likelihood to stall. Krichene [10] used real-time identification of prestall waves to demonstrate active control in laboratory compression systems.

However, Day [11], followed by Camp and Day [12] showed two routes to stall, via modal waves and via "spikes." Their work also shows that prestall waves may not be present prior to stall for the entire operating range of the compressor. In fact, Escuret et al. [13] found no evidence of prestall waves in the high-speed compressor they studied. The spike-type stall was found to be dominant at part speed operation by Hoss et al. [14]. As the available analytical models do not capture all the routes to stall, the model-based control techniques are infeasible for implementation in the production environment. Consequently Bright et al. [15] and Hoss et al. [14] have explored different techniques aimed at nonmodal precursors.

Inoue et al. [16] conducted experiments on a low-speed axial compressor and determined that rotating stall precursors could be detected by the collapse of blade spacing periodicity. They proposed a similarity coefficient as a practical detection parameter. This coefficient utilized data from 200 rotor revolutions and thus was unsuitable for use during rapid transients. Dhingra et al. [17] developed a stability measure based on the unsteady characteristics of the flow field in the tip region of the compressor rotor. Later, Tahara et al. [18] also reported a stall warning index along similar lines.

The present work continues the development of the stability measure reported by Dhingra et al. [17]. This measure, called the correlation measure, has stochastic properties which can be re-

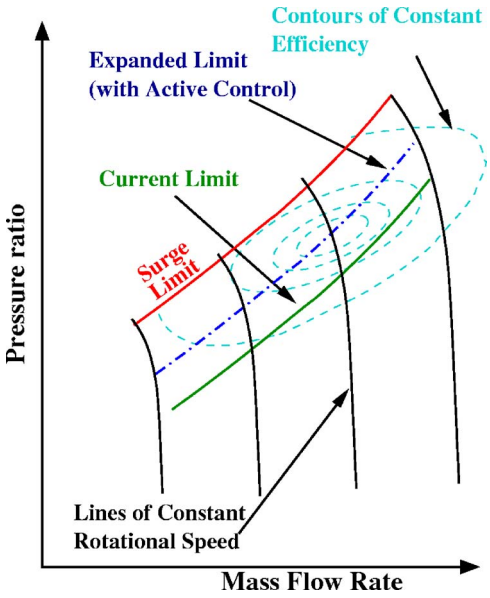


Fig. 1 Expected performance enhancement with active control

lated to the compressor stability margin. An evolving model characterizes this relationship in terms of certain stochastic processes. These processes, part of the *point processes* family, are created due to the intersection of a random process and a constant threshold value. Such intersections are sometimes known as "level crossings." The focus of model development is to gauge the stochastic nature of the correlation measure and thus guide control law design. The model validation utilizes data obtained on three different axial compressors. These include a single-stage low-speed compressor, a four-stage low-speed research compressor, and an advanced, technology demonstrator, six-stage high-speed compressor. This paper presents the details of model development and validation, and concludes with closed loop experimental results to demonstrate the application of correlation measure in compressor stability management.

Experimental Facilities

To establish the fundamental nature of the results presented here, three different compressor systems have been analyzed. While two are low-speed facilities, the third is a modern high-speed compressor. A brief description of the three facilities follows.

GT-Axial Rig. The axial compressor rig located in the School of Aerospace Engineering at Georgia Institute of Technology has been established to facilitate control-oriented studies in surge and rotating stall. It includes an inlet duct, a single-stage axial fan, a compressor discharge duct, a plenum, exhaust duct and a throttle. The low-speed single-stage fan delivers a pressure ratio of about 1.035. It has 14 blades on the rotor and 11 blades on the stator. The design speed is about 11,800 rpm at which the tip Mach number is about 0.3. The plenum is a large metal chamber, capable of withstanding pressures up to 400 psi. It incorporates a self-entraining burner and is thus capable of simulating a combustor. The compressor load is varied by a butterfly valve downstream of the plenum. The compressor is known to have a very abrupt characteristic near the stall inception point. The compressor is instrumented using six high bandwidth pressure sensors. The sensors are mounted flush to the inside of the compressor casing in order to maintain a large overall bandwidth. Data are sampled using a high-performance PC-based data acquisition system.

Low Speed Research Compressor (LSRC). The LSRC is a GE experimental facility that duplicates the relevant aerodynamic features of axial flow compressors in modern gas turbine engines in a large, low-speed machine where very detailed investigations of the flow can be made. Aerodynamic similarity with respect to the Mach number as well as the Reynolds number is used to scale the high-speed airfoils to their low-speed counterparts. This method of testing has been found to be reliable for over 50 years in understanding and designing high pressure compression systems, provided the phenomena being studied are dependent on Reynolds number and not on compressibility.

The LSRC, which has a constant casing diameter of 1.524 m (60.0 in.), was set up with four identical stages in order to simulate the repeating stage environment. The data analyzed here are from the fourth stage. The configuration used is typical of modern engines and has a high hub/tip ratio of 0.85 with low aspect-ratio, high-solidity blading and shrouded stators. The compressor is a low-speed aerodynamic model of the middle block of highly loaded, high-reaction high-pressure compressor, typical of commercial gas turbine engines that are currently in service. Additional information about the LSRC testing technique and the blading is available in Storace et al. [19].

High-Speed Compressor (HSC). The High Speed Compressor (HSC) is a six-stage technology development compressor with high stage loading, higher than that currently found in commercial aircraft engines. The blading designs employed the latest three-dimensional aerodynamics, including forward swept rotors and bowed and swept stators. The compressor rig employs several features designed to model the production engine environment. These include a settling chamber, a goose-neck that simulates the flow path radius change between low-pressure and high-pressure spools, and a volume downstream of the compressor which models a combustor and emulates the surge dynamics of a production environment. A distortion screen at the inlet generates the pressure-only equivalent of pressure and temperature distortions due to typical low-pressure compressors. Finally, an appropriate amount of mass is bled to account for customer as well as turbine cooling bleeds, a part of the production environment.

Correlation Measure

The axial compressor facility at the School of Aerospace Engineering, Georgia Institute of Technology has been used to develop a novel compressor stability management approach. This approach is based on a measure which detects the proximity of the compressor stability limit. The measure is essentially a modified form of autocorrelation function, and is simply referred to as the correlation measure. The standard autocorrelation function for stationary ergodic processes is given by

$$R_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_0^T x(u)x(u+\tau)du \tag{1}$$

where $R(\cdot)$ is the autocorrelation function, x is a sample signal from a random process, and τ is the lag time [20]. The correlation measure is defined on the basis of the repeatability of the pressure time-trace, as observed by a sensor located over the rotor. The pressure time-trace is mostly periodic when the compressor is operating away from the surge line. However, as the boundary of stable operation is approached, this periodicity is progressively disrupted. The loss of periodicity is quantified via the correlation measure, which can be expressed as

$$C(n) = \frac{\sum_{i=n-\text{wnd}}^{n} (p_i p_{i-\text{shaft}})}{\sqrt{\left(\sum_{i=n-\text{wnd}}^{n} p_i^2\right)\left(\sum_{i=n-\text{wnd}}^{n} p_{i-\text{shaft}}^2\right)}} \tag{2}$$

where $C(n)$ is the correlation measure, p is the high-passed pressure signal, i is an index, shaft is the number of samples in one

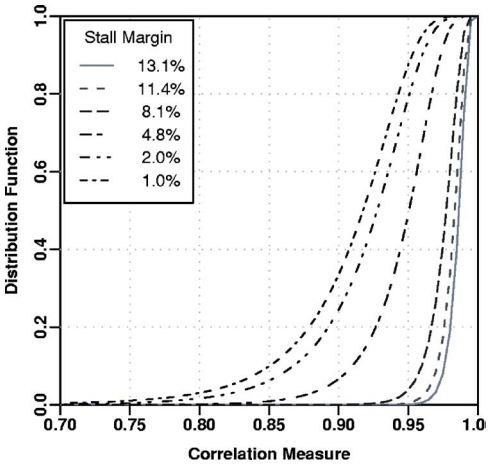


Fig. 2 GT-Axial results: The cumulative distribution of the correlation measure

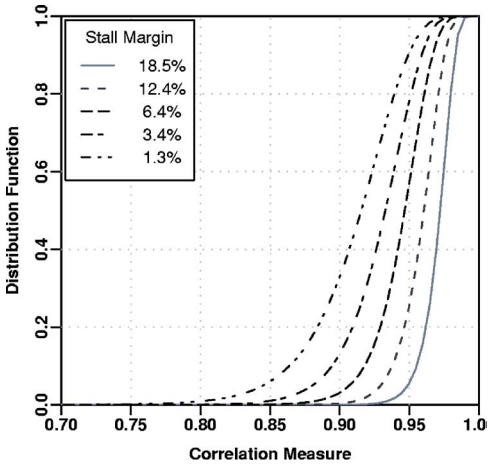


Fig. 3 LSRC results: The cumulative distribution of the correlation measure for compressor at 750 rpm

shaft revolution, n is the current sample index, and wnd is the correlation window in number of samples. The pressure signal is high-passed to ensure that the measured pressure has zero mean. Although wnd can be any value up to the number of samples in one shaft rotation, the experience to date is that a value which spans two to four blades is a good choice. A comparison of Eqs. (1) and (2) shows that the correlation measure is essentially truncated autocorrelation function applied to digitized data, normalized by the respective norms. Consequently, by its very mathematical definition, the correlation measure is bound by 1 from above and −1 from below. When the behavior of the compressor is accounted for, the measure usually stays between 1 and 0, with a value of 1 implying perfect repetition of the pressure trace.

Application to Experimental Data. It has been shown that the average value of correlation measure decreases monotonically with compressor stall margin [17]. Although a convenient metric, average values are not a suitable basis for active control. An average value is slow to respond by definition, and would be unreliable during transients with rapid variations in stall margin. Moreover, the contribution of intermittent loss in blade passage periodicity, which has been the motivation behind the correlation measure formulation, is attenuated by the process of averaging. In the following, the relationship of correlation to compressor stall margin is presented in terms of its distribution function. A distribution function, also known as the cumulative distribution function, can be expressed as

$$F(x) = P(X < x) \tag{3}$$

where X is a random variable, x is a specific value of the random variable, and $P(\cdot)$ is the probability. As an example, if a given sample of correlation measure spends 20% of its time below 0.8, the corresponding value of its distribution function, $F(0.8)$, would equal 0.2.

The distribution function of the correlation measure for a specific data set obtained on the GT-Axial facility is shown in Fig. 2. Each curve corresponds to steady state operation of the compressor and is characterized by the compressor stall margin. It can be observed that the distribution curves shift left, signifying that the mean value of the correlation measure is decreasing with decreasing stall margin. Further, the decreasing slope, or alternatively the larger difference between the minimum and maximum values of correlation measure, conveys an increase in variance about the mean, as the stall margin is decreased. Note that the minimum value of the correlation measure is given by the x-intercept in Fig. 2. For example, while the correlation varies from about 0.925 to 0.99 in the 8.1% stall margin case, it varies from 0.7 to 0.975 for the 2.0% case. It may also be noted that the different distribution curves appear to be of the same family. However, the existence or form of this family has not been identified to date.

Further inspection of Fig. 2 reveals an important feature of the correlation measure. The difference between the different stall margin cases is very pronounced in terms of the time spent by the correlation measure below a fixed value. For example, in Fig. 2 if a value of 0.9 is considered, correlation measure never falls below it for the least loaded case (13.1% stall margin). On the other hand, it roughly has a 0.35 probability of being lower than 0.9 for the most loaded case (1.0% stall margin).

Similar analysis has been performed using data obtained on the GE-LSRC facility. The results, presented in Figs. 3 and 4, show a consistent trend across the different speed lines. This trend, also observed in the GT-Axial results (Fig. 2), shows that the mean value of the correlation measure decreases while the variance about the mean increases as the stall margin is decreased. Although the qualitative behavior is the same, the quantitative values are dependent on the speed. In particular, the correlation measure exhibits a smaller change for the same change in compressor stall margin when the compressor is operating at 800 rpm, as compared to the 750 rpm case. This difference could, in part, be attributed to differences in the stage load distribution for the two speeds. However, it is possible that correlation measure has some dependence on compressor speed, which would need to be ad-

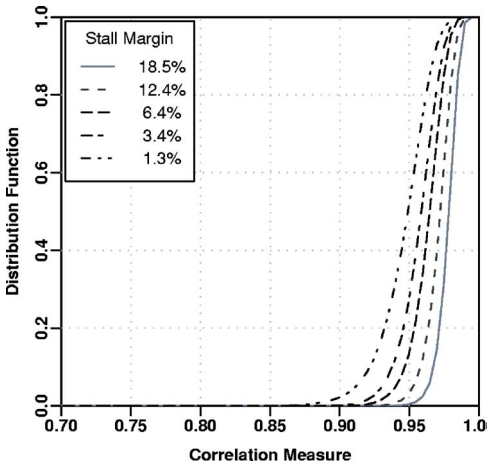


Fig. 4 LSRC results: The cumulative distribution of the correlation measure for compressor at 800 rpm

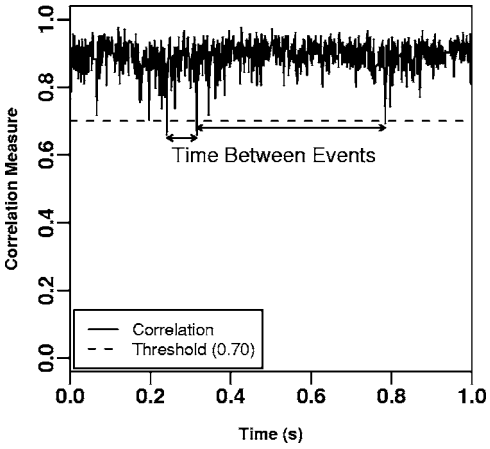


Fig. 5 Random occurrence of events. An event is defined as downwards crossing of a threshold.

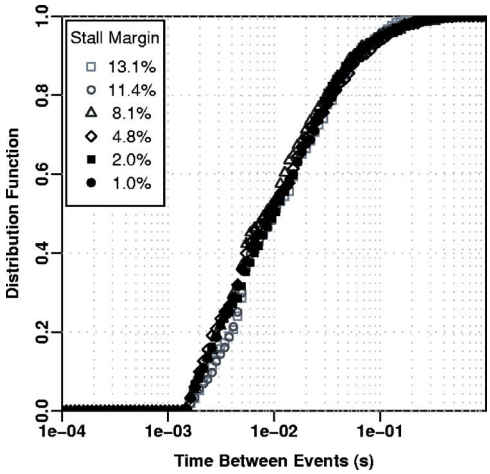


Fig. 6 GT-Axial results: The cumulative distribution of TBE. The threshold is set at the average correlation measure for each case.

dressed in active stability management applications.

Although the specific results have been omitted here, similar relationship of the correlation measure to the compressor stall margin has been observed using data obtained on the HSC facility.

Stochastic Model

As mentioned earlier, average value of the correlation measure is not a good basis for active stability management. An inspection of the experimentally obtained correlation measure shows that along with a change in the average value, the correlation measure exhibits a tendency to dip from time to time. These dips increase in frequency and magnitude as the compressor approaches its limit of stable operation. The appropriate approach, then, is to monitor the dips in the correlation measure and generate an alarm event whenever the measure falls below a specified threshold. Recall that the results pertaining to the cumulative distribution function of the correlation measure also suggest that time spent by it, below a given threshold, could yield a suitable criterion for detection of compressor stability limit. The basic idea of events is illustrated in Fig. 5, which shows the time trace of the correlation measure calculated using data from the GT-Axial facility. The apparent randomness of the dips in the correlation measure leads to the alarm events that are randomly distributed in time. In a possible implementation, a controller would react to these alarm events by initiating suitable preventive actuation. Consequently, the time between two consecutive events (TBE), as identified in Fig. 5, is an important metric. A stochastic model is hence sought which captures the distribution of the events and its relationship to the stall margin.

The random occurrence of the events makes them a candidate *point process*. Point processes are a class of stochastic processes where events with no duration occur successively in time, in a more or less random manner. An everyday example of a point process is the arrival of customers in a bank. A Poisson process is the simplest possible point process, and serves as a model for events such as radioactive disintegrations. Motivated by the Poisson process, along with experimental observations, it is proposed that TBE is exponentially distributed, and its distribution function is given by

$$F(\delta_t) = 1 - e^{-\mu\delta_t} \tag{4}$$

with the parameter μ defined as

$$\mu = \lim_{T\to\infty} \frac{N(0,T)}{T} \tag{5}$$

where δ_t is the time between consecutive events, and N is the total number of events in time period $(0,T)$. Thus, the parameter μ is the average number of events. In practice, $(0,T)$ is a finite period of time, large enough for the average to be statistically meaningful.

Although events in a Poisson process have infinitesimal duration, in practice, processes where the duration of events is much smaller than the time between successive events can sometimes be approximated as Poisson processes. In such a case, there would be a limit on the event rate for the approximation to be valid. In the present case, as the downward crossing of a specified threshold is designated as an event, the events indeed have no duration, and thus there are no theoretical constraints on event rates. However, it is important to reiterate that the Poisson process model is still an approximation of the observed behavior.

Model Validation

GT-Axial Rig. In this section, the validity of the proposed stochastic model is evaluated via analysis of sets of data acquired on the GT-Axial Rig, a laboratory axial compressor facility. As part of each set, pressure data from a single transducer are acquired for different steady operating conditions characterized by their stall margins.

To illustrate the uniqueness of the underlying process that characterizes the correlation measure, and its independence of the compressor steady state, the statistics of the time between events have been analyzed. As mentioned earlier, an event is defined as the downward crossing of a given level, C_{th} by the correlation measure $C(n)$. A specific example of the distribution function of time between events (TBE) is presented in Fig. 6. Each curve corresponds to steady state operation of the compressor at a different stall margin. For each stall margin, the threshold level C_{th} is set equal to the mean value of the corresponding $C(n)$, and is thus a function of the stall margin. For this special choice of threshold, all the TBE distributions collapse to a common function, irrespective of the stall margins. Consequently it may be inferred that the variation of $C(n)$ about its mean value, under steady state conditions, is driven by an invariant process.

If the threshold level C_{th} is kept fixed across all runs, a more pertinent choice, the situation is different. The results in Fig. 7 correspond to a fixed threshold level of 0.76. No event is observed for stall margins higher than 4.8%. As the compressor is moved closer to its surge limit, the margin reduces from 4.8% to 1.0%, the probability of an event occurring in a given time increases. As an example, with 1.0% stall margin, the probability of the correlation measure dipping below 0.76 over 100 ms interval is

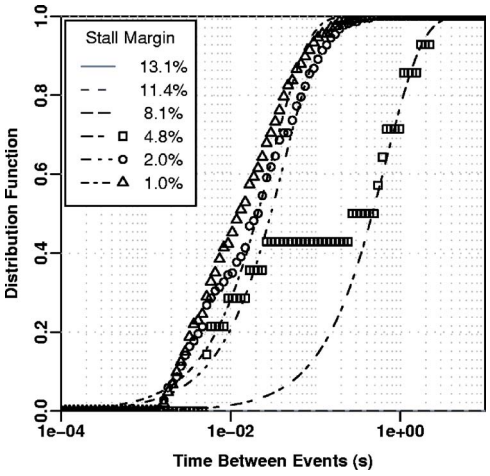


Fig. 7 GT-Axial results: The cumulative distribution of TBE with C_{th} of 0.76. The experimentally observed distribution (markers) is compared to that predicted by the proposed model (solid lines).

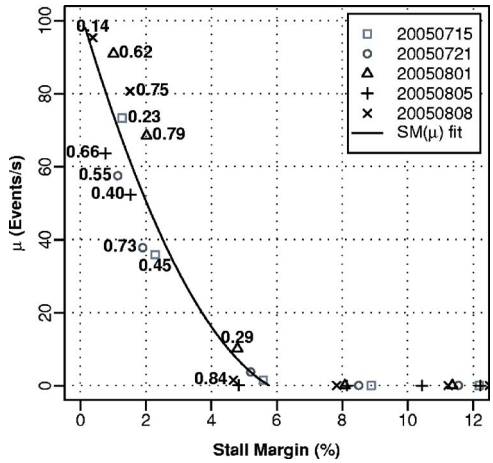


Fig. 9 GT-Axial results: The day to day variation of the observed average number of events. The labels next to each marker designate the error in stall margin as estimated via least-square error curve fit.

roughly 0.9 whereas the same for the 4.8% case is close to 0.4, based on the observed distribution of events. This is a consequence of an increased number of events in a given time period.

The fairly close agreement of the predicted and the experimentally obtained curves, as observed in Fig. 7, suggests that the proposed model is valid for the problem at hand. Similar analysis for different threshold levels in this as well as other sets of data shows that the TBE distribution for lower stall margins is consistently captured by the model. On the other hand, for the larger stall margins some discrepancy between the two has been observed. In part, this discrepancy could be due to the low number of events actually observed at high stall margins, reducing the validity of statistical analysis. Given the limited size of data available, combined with the low event rates, a smaller number of events would necessarily be observed at high-stall margin operation of the compressor.

The stochastic model is explicitly dependent on the average number of events, i.e., the parameter μ. Consequently, data from GT-Axial facility has been used to illustrate the behavior of this parameter and the results are presented in Fig. 8. The figure shows variations in μ with stall margin for a variety of threshold values, C_{th}. The results show that for a given C_{th}, μ increases with decreasing stall margin. In other words, events occur at a higher rate at lower stall margins. On the other hand, μ decreases as the threshold C_{th} is lowered, i.e., fewer events are observed, at a given stall margin. Events cannot be observed for the higher stall margins. More significantly, the first appearance of an event shifts to lower stall margins as C_{th} is lowered. For example, some events can be observed at 4.8% stall margin when C_{th} is set at 0.76. If it is lowered to 0.64, events only appear when the compressor stall margin is reduced to 2%. Thus even though dips in correlation measure are present at all stall margins, their magnitude is larger for lower stall margin.

In order to investigate the repeatability of these results, average number of events for different stall margins have been calculated using five different sets of data. These data sets were acquired on different days with varying ambient conditions, spread over three weeks. The results are presented in Fig. 9 and exhibit good consistency for the cases studied. In this figure, the markers represent observed μ values as a function of stall margin, and are grouped according to the different data sets. A least square fit has been used to first obtain μ as a function of stall margin (SM). The one-to-one part of this curve has then been inverted to yield a function of the form SM(μ). The labels next to each marker identify the horizontal distance between it and the SM(μ) curve

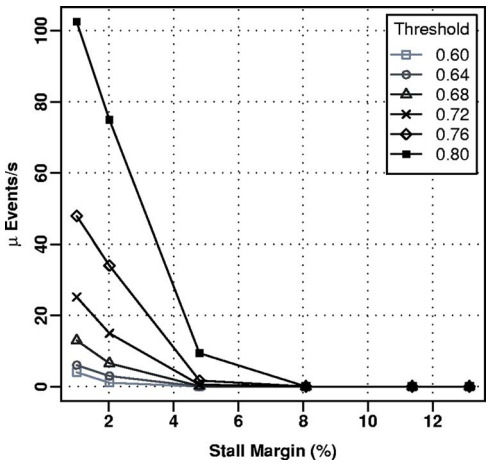


Fig. 8 Variation of the average number of events with stall margin. The threshold value plays a dominant role in this relationship.

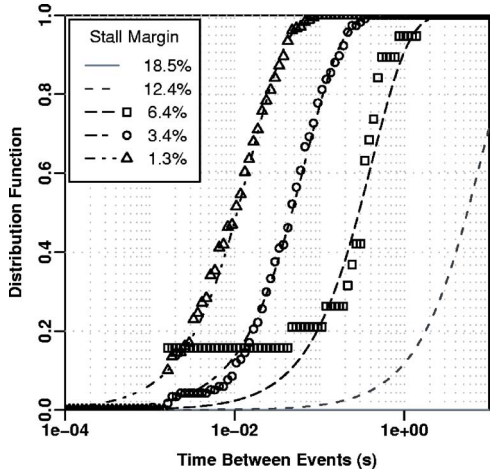


Fig. 10 LSRC results: The cumulative distribution of TBE for compressor at 750 rpm. The threshold C_{th} is set 0.86.

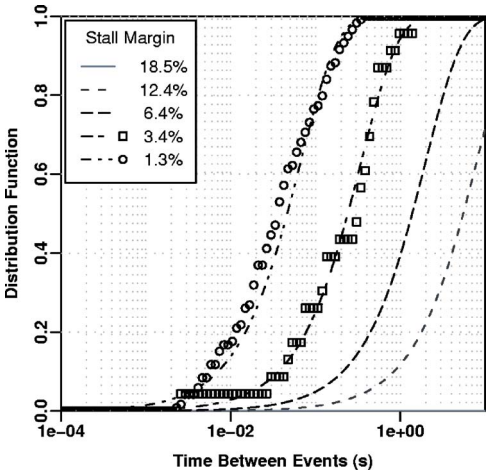


Fig. 11 LSRC results: The cumulative distribution of TBE for compressor at 800 rpm. The threshold C_{th} is set 0.92.

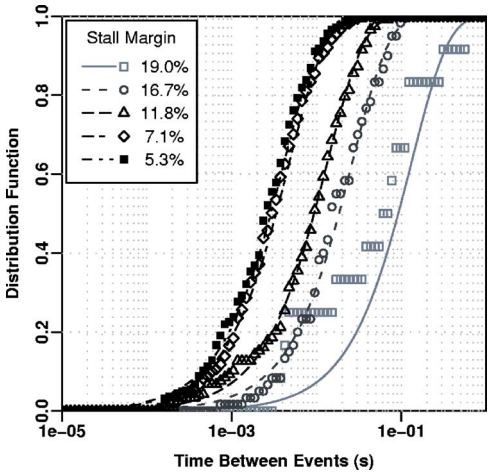


Fig. 13 HSC results: The cumulative distribution of TBE for the second stage. The threshold C_{th} is set 0.62.

shown. This distance, in essence, is the error in stall margin that would be predicted through observed μ. For example, about 38 events/s have been observed for the 2% stall margin case, part of the 20050721 data set. On the basis of the fitted SM(μ) curve, 38 events/s would correspond to about 2.7% SM, leading to an error of about 0.7% SM (actual error 0.73% SM). The nature of SM(μ) is such that higher stall margins cannot be distinguished from each other.

Figure 9 emphasizes an important aspect of the correlation measure technique. Tracking events generated by correlation measure can only be used to detect the proximity of the compressor stability limit. The correlation measure based approach is not applicable as a general stall margin estimator. However, when operating close to the surge line, this approach enables a real-time estimation of compressor stall margin with reasonable accuracy.

Low-Speed Research Compressor. The proposed stochastic model is next validated against data from the GE-LSRC facility. Two sets of data, each with the compressor operating at different speed, have been analyzed. The compressor loading is varied in a quasi-steady manner via a discharge valve. These results are from a single sensor located roughly over the mid-chord of the fourth rotor.

The distribution of correlation measure, while qualitatively similar, was found to be dependent on the compressor speed (Figs. 3 and 4). On the other hand, the occurrence of the alarm events is less sensitive to the compressor operating speed. This is illustrated in Figs. 10 and 11. In either case, a very similar distribution of events is observed. The experimentally observed distribution is shown via markers, while the prediction is depicted by curves. For higher stall margin cases, namely 18.5% and 12.4%, too few events were observed and the experimental cumulative distribution could not be determined. Overall, the stochastic model does a reasonable job of predicting the distribution of the TBE.

High-Speed Compressor. The HSC facility has been used to evaluate the applicability of correlation measure and the corresponding stochastic model to a modern high-speed compressor. Moreover, the availability of pressure data from multiple stages extends the validation of the proposed model. The available data sets contain pressure signals from three stages and the results presented here are based on one sensor per stage. The sensor is located over the rotor midchord in each case.

Figures 12, 13, and 14 contain results for stages 1, 2 and 3, respectively. Although these results are for three specific threshold values, analogous results have also been obtained for other values. As in the previous two compressors, the stochastic model is able to capture the distribution of the TBE quite well. There is a noticeable deficiency at lower TBEs for the first stage, as observed in Fig. 12, which is recovered at the higher TBEs.

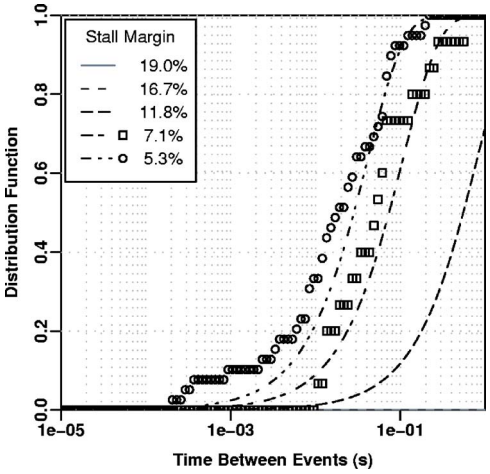


Fig. 12 HSC results: The cumulative distribution of TBE for first stage. The threshold C_{th} is set 0.82.

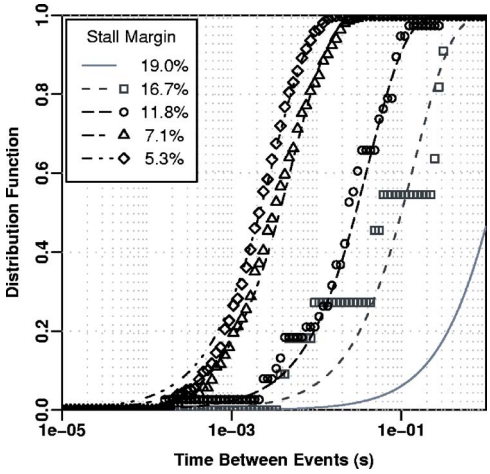


Fig. 14 HSC results: The cumulative distribution of TBE for the third stage. The threshold C_{th} is set 0.56.

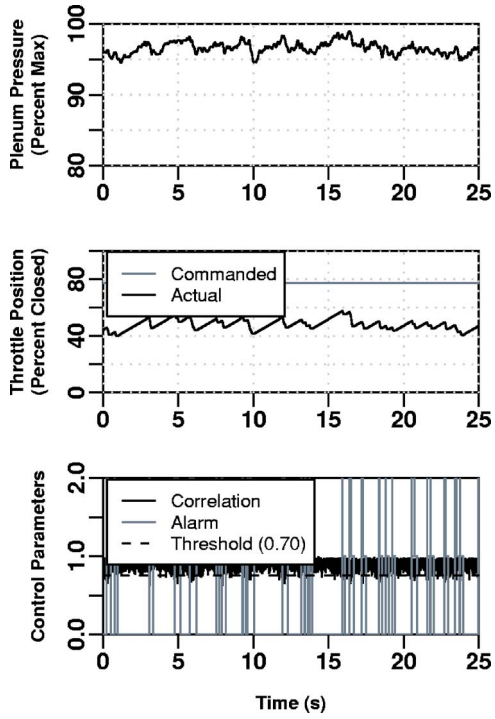


Fig. 15 GT-Axial rig results. The simple open-pause-close control law keeps the system close to its maximum pressure operation.

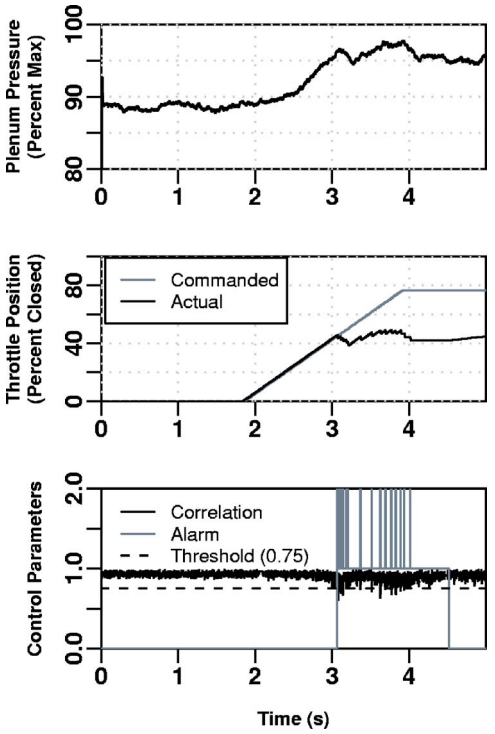


Fig. 16 GT-Axial rig results. The open-pause-close controller is able to avoid surge under throttle transients.

Experimental Demonstration

The GT-Axial compressor facility has been used to demonstrate correlation measure based surge avoidance scheme. In the results presented here, a very simple control scheme has been utilized. A common strategy is employed for two cases: Ride the limit, and surge line avoidance under fast transients. An alarm event is generated when the correlation measure $C(n)$ crosses a selectable threshold level. An open-throttle control action is triggered by an alarm event. The controller waits for a fixed short interval of time, after which the throttle is closed at a *fixed rate*, up to its operator commanded value. This open-pause-close strategy is followed for each observed event. The objective for the "ride the limit" case is to keep the plenum pressure close to the maximum deliverable by the compressor, without crossing the stall boundary. This simple control law performs adequately in this case and can hold the system close to the stall line indefinitely. However, for the sake of clarity, Fig. 15 only shows results for a 25 s segment. The system is held at about 4% to 5% stall margin by the controller. The operator commanded value is deliberately set deep into the stall region, ensuring that any deficiency of the control method will be manifested as a surge in the system. Further, the intermittent nature of the alarms is evident in the lower chart of Fig. 15. As noted, the threshold level for the correlation measure has been set at 0.70 in this case.

The "fast transient" case is a ramp command which closes the throttle from its fully open position to one corresponding to deep stall in 2 s. As a matter of fact, the system reaches the stall line in about 1 s into the transient. This case is intended to emulate transients analogous to those encountered in transient gas turbine operation. Consequently, the objective is to prevent the system from surging while following the operator commanded value as closely as possible. A typical run, with threshold level set at 0.75, is shown in Fig. 16. As can be observed in these results, the controller prevents the system from crossing the surge line. Similar to the "ride the limit" scenario, with continuous active adjustment of the throttle, the controller tries to achieve the maximum pressure deliverable by the system.

The emulation is limited in scope as it only entails a rapid variation of stall margin. Significantly, unlike gas turbine transients, there is no change in the compressor rotating speed. However, the feasibility of correlation measure based stability management has been demonstrated in full-scale engine tests conducted by GE Aircraft Engines [21].

In order to emphasize the stochastic nature of the correlation measure based alarms, a sequence of ramp transients has been performed. The results are summarized in Fig. 17. Although the figure contains roughly ten transients, active surge avoidance

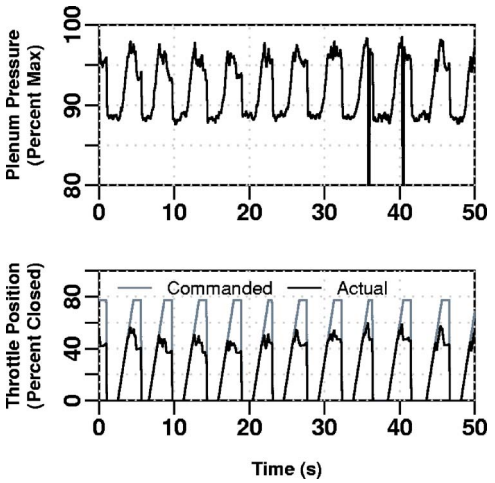


Fig. 17 Importance of threshold parameter C_{th} for surge limit avoidance. A poor choice coupled with stochastic nature of alarms can lead to failure of the controller.

failed in two out of the twenty transients performed. This translates into a 90% success rate at the selected threshold level of 0.75. This value was *deliberately* selected to showcase the importance of the threshold level in active surge avoidance. The success rate is a function of the threshold value and can be improved by adjusting the said level.

Conclusions

The correlation measure attempts to translate the unsteady characteristics of the flow field over the compressor rotor in to a stability measure. In the past, this measure has been shown to decrease monotonically with stall margin. The present work provides a model for the time between successive events, where an event is defined as the downward crossing of a prespecified threshold. A potential control scheme would react to these events, and consequently their arrival rate is an important variable. The largest time between events represents the worst-case scenarios from a control system perspective. The proposed stochastic model predicts the distribution of event occurrence based on the average number of events observed. In essence, the model predicts that the events are exponentially distributed in time. This stochastic model has been validated with data obtained on three very different axial compressor designs. These include a single-stage low-speed axial compressor, a four-stage low-speed research compressor, and an advanced technology demonstrator high-speed compressor. It has been found that in all three cases, the number of events increases with decreasing stall margin. At the same time, the maximal value of the time between two successive events decreases, a fact captured by the stochastic model.

The key ideas associated with the proposed approach to compressor control have been experimentally verified. This validation is in the form of demonstration of closed-loop control on the GT-Axial Rig, a laboratory axial compressor facility. The experimental demonstration so far has used relatively simple control laws. The validity of the "proximity to the limit" detection method has been established. This enables the focus to be shifted to the development of integrated control laws.

Acknowledgment

This study was conducted under the NASA URETI on Aeropropulsion and Power Technology (UAPT) and the GE Aircraft Engines University Strategic Alliance (GEUSA) program at the Georgia Institute of Technology.

Nomenclature

$C(n)$ = correlation measure
C_{th} = threshold value for the correlation measure
TBE = time between successive events
i = index
p = pressure signal, high-passed
shaft = number of samples in one shaft revolutions
n = current sample index
wnd = correlation window in number of samples
μ = average number of events

References

[1] Moore, F. K., and Greitzer, E. M., 1986, "Theory of Post Stall Transients in Axial Compression Systems—Part I: Development of Equations," Trans. ASME: J. Eng. Gas Turbines Power, **108**(1), pp. 68–76.
[2] Moore, F. K., and Greitzer, E.M., 1986, "Theory of Post Stall Transients in Axial Compression Systems—Part II: Application," Trans. ASME: J. Eng. Gas Turbines Power, **108**(2), pp. 231–239.
[3] Epstein, A. H., Williams, J. E. F., and Greitzer, E. M., 1989, "Active Suppression of Aerodynamic Instabilities in Turbomachines," J. Propul. Power, **5**(2), pp. 204–211.
[4] Weigl, H. J., Paduano, J. D., Frechette, L. G., Epstein, A. H., Greitzer, E. M., Bright, M. M., and Strazisar, A. J., 1998, "Active Stabilization of Rotating Stall and Surge in a Transonic Single-Stage Axial Compressor," J. Turbomach., **120**, pp. 625–636.
[5] Badmus, O. O., Chowdhury, S., Eveker, K. M., Nett, C. N., and Rivera, C. J., 1995, "Simplified Approach for Control of Rotating Stall—Part 1: Theoretical Development," J. Propul. Power, **11**(6), pp. 1195–1209.
[6] Badmus, O. O., Chowdhury, S., Eveker, K. M., Nett, C. N., Rivera, C. J., 1995, "Simplified Approach for Control of Rotating Stall—Part 2: Experimental Results," J. Propul. Power, **11**(6), pp. 1210–1223.
[7] Gysling, D. L., and Greitzer, E. M., 1994, "Dynamic Control of Rotating Stall in Axial Flow Compressors Using Aeromechanical Feedback," ASME Paper No. 94-GT-292.
[8] D'Andrea, R., Behnken, R. L., and Murray, R. M., 1997, "Rotating Stall Control of an Axial Flow Compressor Using Pulsed Air Injection," J. Turbomach., **119**(4), pp. 742–752.
[9] Tryfonidis, M., Etchvers, O., Paduano, J. D., Epstein, A. H., and Hendricks, G. J., 1995, "Prestall Behaviour of Several High-Speed Compressors," J. Turbomach., **117**, pp. 62–80.
[10] Krichene, A., 2001, "Active Identification and Control of Aerodynamic Instabilities in Axial and Centrifugal Compressors," Ph.D. thesis, Georgia Institute of Technology, Atlanta.
[11] Day, I. J., 1993, "Stall Inception in Axial Flow Compressors," J. Turbomach., **115**, pp. 1–9.
[12] Camp, T. R., and Day, I. J., 1997, "A Study of Spike and Modal Stall Phenomena in a Low Speed Axial Compressor," ASME Paper No. 97-GT-526.
[13] Escuret, J. F., and Garnier, V., 1996, "Stall Inception Measurements in a High-Speed Multistage Compressor," J. Turbomach., **118**, pp. 690–696.
[14] Hoss, B., Leinhos, D., and Fottner, L., 2000, "Stall Inception in the Compressor System of a Turbofan Engine," J. Turbomach., **122**, pp. 32–43.
[15] Bright, M. M., Qammar, H. K., Weigl, H. J., and Paduano, J. D., 1997, "Stall Precursor Identification in High-Speed Compressor Stages Using Chaotic Time Series Analysis Methods," J. Turbomach., **119**(3), pp. 491–499.
[16] Inoue, M., Kuroumaru, M., Iwamoto, T., and Ando, Y., 1991, "Detection of a Rotating Stall Precursor in Isolated Axial Flow Compressor Rotors," J. Turbomach., **113**(2), pp. 281–289.
[17] Dhingra, M., Neumeier, Y., Prasad, J. V. R., and Shin, H., 2003, "Stall and Surge Precursors in Axial Compressors," AIAA Paper No. 2003–4425.
[18] Tahara, N., Outa, E., Kurosaki, M., Nakakita, T., Ohta, Y., and Tsurumi, Y., 2004, "Early Stall Warning Technique for Axial Flow Compressors," *Proceedings of the ASME Turbo Expo 2004*, ASME, New York, Vol. 5A, pp. 375–384.
[19] Storace, A. F., Wisler, D. C., Shin, H.-W., Breacher, B. F., Ehrich, F., Spakovsky, Z. S., and Martinez-Sanchez, M., 2000, "Unsteady Flow and Whirl-Inducing Forces in Axial-Flow Compressors—Part I: Experiment," ASME Paper No. 2000-GT-0564.
[20] Bendat, J. S., and Piersol, A. G., 1986, *Random Data: Analysis and Measurement Procedures*, 2nd ed., Wiley, New York.
[21] Christensen, D., Cantin, P., Gutz, D., Szucs, P., Wadia, A., Armor, J., Dhingra, M., Neumeier, Y., and Prasad, J., 2006, "Development and Demonstration of a Stability Management System for Gas Turbine Engines," ASME Paper No. GT-2006-90324.

Performance of Gas Turbine Power Plants Controlled by the Multiagent Scheme

Lorenzo Dambrosio
DIMeG Sez. Macchine ed Energetica,
Politecnico di Bari,
Via Re David 200,
70125 Bari, Italy
e-mail: dambrosio@poliba.it

Marco Mastrovito
DIASS,
Polytechnic University of Bari,
Viale del Turismo 8,
74100 Taranto, Italy
e-mail: m.mastrovito@poliba.it

Sergio M. Camporeale
DIMeG Sez. Macchine ed Energetica,
Politecnico di Bari,
Via Re David 200,
70125 Bari, Italy
e-mail: camporeale@poliba.it

In recent years the idea of artificial intelligence has been focused around the concept of rational agent. An agent is an (software or hardware) entity that can receive signals from the environment and act upon that environment through output signals, trying to carry out an appropriate task. Seldom agents are considered as stand-alone systems; on the contrary, their main strength can be found in the interaction with other agents, constituting the so-called multiagent system. In the present work, a multiagent system was chosen as a control system of a single-shaft heavy-duty gas turbine in the multi input multi output mode. The shaft rotational speed (power frequency) and stack temperature (related to the overall gas turbine efficiency) represent the controlled variables; on the other hand, the fuel mass flow (VCE) and the variable inlet guide vanes (VIGV) have been chosen as manipulating variables. The results show that the multiagent approach to the control problem effectively counteracts the load reduction (including the load rejection condition) with limited overshoot in the controlled variables (as other control algorithms do) while showing a good level of adaptivity, readiness, precision, robustness, and stability. [DOI: 10.1115/1.2718567]

Introduction

The dynamic behavior of a gas turbine is inherently nonlinear and characterized by a large number of internal parameters. The design of the control system is often based on a linearized model that is generally applicable only in a limited range around such a steady state condition [1,2]. Typically the linearization of the system is carried out for various operating points and the control parameters are determined for these operating points using, e.g., linear quadratic regulator theory. Nonlinear control algorithms are more suitable to control nonlinear systems In fact, for a gas turbine power plant a number of such algorithms have been proposed. A few of them can be divided in two main branches: nondeterministic and knowledge-based. The one step ahead adaptive and the weighted one step ahead adaptive (Refs. [3–5]) algorithms refer to the nondeterministic family. Such algorithms employ the least squares algorithm parameter estimator in order to estimate the coefficients of a DARMA model of the controlled system. In particular, the one step ahead controller determines the control law from the DARMA model, trying to annihilate the control error one step in the future. On the other hand, the weighted one step ahead adaptive control algorithm considers a penalty associated with the control effort by means of an appropriate cost function (Refs. [5,6]). In Ref. [7] a nonlinear control technique for the gas turbine based on self-tuning control parameters is described. Finally, Refs. [8,9] deal with model reference predictive control algorithms.

As far as the knowledge-based methods are concerned, they usually are based either on fuzzy logic system and artificial neural networks structure, or on the model reference approach. The first ones employ sets of fuzzy rules (supplied by expert or by numerical/experimental data) that perform a mapping between input and output variables (Refs. [10,11]). The second method needs to be trained (by means of input-output samples) in order to operate the input-output mapping (Ref. [12]). Finally, the latter employs a model reference of the controlled system in order to evaluate the appropriate control law (Refs. [13,14]).

All the foregoing algorithms present, in a different degree, good characteristics in terms of adaptivity readiness, precision, robustness, stability, etc. Nevertheless, it is arduous that a centralized single algorithm shows good performance for all the mentioned features. Usually a compromise choice has to be made. A way to overcome such limitations is represented by the multiagent approach. Recently the idea of artificial intelligence has been focused around the concept of rational agent. An agent is an (software or hardware) entity that can receive signals from the environment and act upon that environment through output signals. In general an agent always tries to carry out an appropriate task; such an agent is called a rational agent. Generally agents are not considered as stand-alone systems. Their main strength can be found in the interaction with other agents in several different ways. Therefore a system composed of a group of agents that can potentially interact with each other and the environment in order to fulfill a specific task is called a multiagent system (MAS) (Refs. [15,16]).

The MAS approach has been used to manage the control process of a single-shaft heavy-duty gas turbine for electric power application. In this work, the MAS control scheme has been applied in multi input multi output (MIMO) mode for power regulation, assuming as controlled variables, the rotational speed (which is closely related to the power frequency), and by the stack temperature (which affects the overall efficiency) while using as control variables the fuel rate and the angle of the variable inlet guide vanes (VIGV), the plant being assumed to undergo sudden variations of the electric load.

The agents employed in the proposed MAS are based on a fuzzy logic scheme. Such a choice allowed the authors, on the one hand, to deal with agents featured by an adequate level of flexibility, thanks to the fuzzy logic nature of each agent; on the other hand, the fuzzy logic system is able to handle approximate data and data affected by uncertainties. Moreover, since each agent has to face the entire control problem, it has been possible to maintain a low level of complexity for the fuzzy logic structure of each agent.

In the Results, it will be shown that the MAS approach to the

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received September 20, 2006; final manuscript received December 15, 2006. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2006: Land, Sea and Air (GT2006), May 8–11, 2006, Barcelona, Spain, Paper No. GT2006-90590.

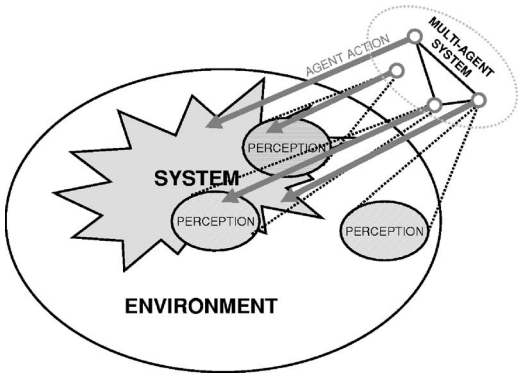


Fig. 1 Multiagent system layout

control problem, applied to the single-shaft heavy-duty gas turbine effectively counteracts the load reduction (even during the load rejection condition) with limited overshoot in the controlled variables (as other control algorithms do) while showing good level adaptivity readiness, precision, robustness, and stability.

Multiagent System

Agents constituting a MAS present a number of features that affect the behavior of the MAS. Such features deal with agent design, environment, perception, control, and knowledge. For a complete description of all the agent's characteristics refer to Refs. [15,16].

In general, a MAS performs a task that is more complex than the subtasks fulfilled by single agents. In a MAS there is not a central process that collects information from each agent and then decides what action each agent should take. In a cooperative or team MAS, distributed decision making results in asynchronous computation and certain speedups, but it also has the downside that appropriate coordination mechanisms need to be additionally examined. The MAS aims at decomposing the problem, allocating subtasks to agents, and synthesizing partial results, in order to:

- Reduce the computational effort due to the asynchronous and parallel computation.
- To obtain a more robust control system, in the sense that the whole system can undergo a "graceful degradation" when one or more agents fail.
- To increase scalability and flexibility, since it is easy to add new agents to the system.
- To maintain a simple architecture for each agent.
- To enable the agent to communicate one to each other in order to improve the ensamble of the action on the system.

As stated earlier, an agent is anything that can be viewed as perceiving its environment through input signal and acting upon that environment through output signal (Fig. 1). Each agent has to accomplish such an action performing an appropriate task. The performance measure is typically defined by the user (the designer of the agent) and reflects what the user expects from the agent in the task at hand. In other words, an agent has to face a decision-making problem. Such an aspect is performed by the so-called *policy function* that associates a particular agent action to a given perception condition. In the present paper, the authors defined the policy function as a fuzzy logic system. The reasons for such a choice lie on the two following aspects:

- to deal with agents featured by an adequate level of flexibility, so that the agent learns how to fulfill the assigned task by a set of rules;
- the fuzzy logic system, thanks to the filtering features of the fuzzification process, will able to handle both quantitative and qualitative rules.

Moreover, the decentralized MAS is able to maintain a low level of complexity for the fuzzy logic structure of each agent, since each agent has not to face the entire optimization problem.

In this context, the perception is considered by the fuzzy logic system as the input vector whereas the fuzzy logic output represents the agent action. In the following section a concise description of the fuzzy logic system will be outlined.

Fuzzy Logic System

The fuzzy logic system (FLS) employs a set of N fuzzy linguistic rules. These rules may be provided by experts or can be extracted from numerical data. In either cases, such rules are expressed as a collection of if-then statements. Therefore, a fuzzy rule base, R, containing N fuzzy rules can be expressed as

$$R = [\text{Rule}_1, \ldots, \text{Rule}_j, \ldots, \text{Rule}_N]$$
$$\text{Rule}_j\text{: if } [\mathbf{z} \text{ is } \tilde{\mathbf{A}}], \quad \text{then } [\mathbf{u}(k) \text{ is } \boldsymbol{\beta}_i] \tag{1}$$

where k refers to the iterative index variable and $\mathbf{z}=[z_1, \ldots, z_l]^T$ are all of the l fuzzy inputs to the FLS (that coincide with the agents perception). On the other hand, $\mathbf{u}(k)=[u_1(k), \ldots, u_m(k)]^T$ and $\beta_i=[\beta_i^1, \ldots, \beta_i^m]^T$ are the FLS fuzzy output (agent output) and the consequent fuzzy set, respectively. In the antecedent of rule Rule_i, the term $\tilde{\mathbf{A}}=[\tilde{A}_i^1, \ldots, \tilde{A}_i^l]^T$ represents the vector of the fuzzy sets referring to the input fuzzy vector $\mathbf{z}$. The membership functions of both the antecedent and consequent, $\tilde{\mathbf{A}}$ and β, respectively, have been chosen to be Gaussian. Such a choice allows an easier handling of the rule composition inference engine; moreover the Gaussian membership function can be easily tuned in order to set up appropriate fuzzy partitions for each input domains. The inference engine employs a *product inference* for the rule implication. Finally the fuzzification and defuzzification processes have been carried out by means of a Gaussian fuzzifier and a *height* defuzzifier, respectively. Gaussian fuzzifier, in particular, acts as a filter for the input uncertainties. As a concluding remark, the resulting fuzzy structure based on Gaussian fuzzifier, Gaussian membership functions of both antecedents and consequents, product inference, and height defuzzifier represents a generic fuzzy frame which is suitable to be applied to different agents performing different tasks. For a detailed description of the fuzzy logic system implementation, please refer to Ref. [17].

Agent Description

In this section a brief description of the agents employed to manage the control process of a single-shaft heavy-duty gas turbine will be outlined. In particular, the MAS control scheme has been applied in the MIMO mode, where the controlled variables are represented by the shaft rotational speed (power frequency) and stack temperature (related to the overall gas turbine effi-

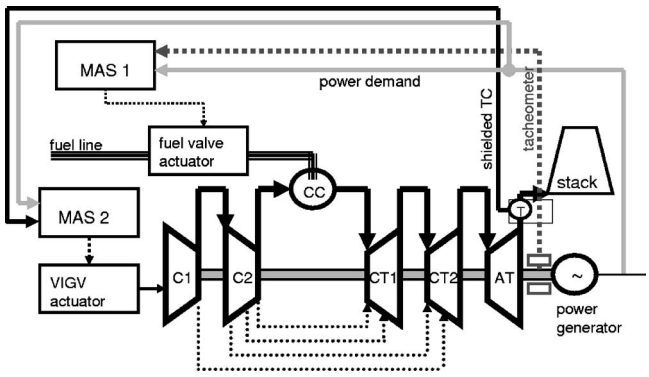


Fig. 2 Gas turbine control signal layout

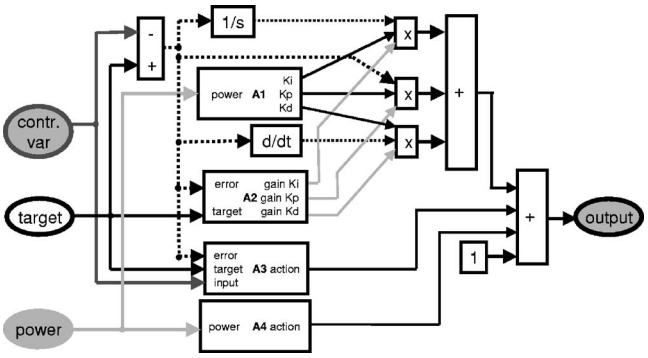


Fig. 3 Multiagent scheme

ciency); on the other hand, the fuel mass flow (VCE) and the variable inlet guide vanes (VIGV) have be chosen as control variables.

As illustrated in Fig. 2, the MIMO control problem has been faced by means of two independent MAS: MAS1 and MAS2, for the rotational speed and stack temperature, respectively. Both the MASs share the same internal structure (Fig. 3) composed of four agents based on a fuzzy policy function. For such reason, in the following, a description of each agent composing the generic MAS will be outlined, omitting any direct reference to the shaft rotational speed and/or stack temperature control problem for the sake conciseness.

Starting from classical control applications which adopt ordinary (or scheduled) PID controllers, the first agent, called AGENT1, performs a fuzzy selection of the PID parameters according to the requested electrical power. In other words, considering a scheduled PID controller, whose parameters have been optimized, by using the Ziegler-Nichols method, for different load conditions, the AGENT1 operates a fuzzy mapping between the requested electrical load and the PID parameters. The PID parameters employed into the AGENT1 base of rules have been optimized taking into account, not only the controller promptness and accuracy, but also the rate of approach of controlled variable to the reference value. In this perspective, the perception for the AGENT1 is represented by electrical power demand, whereas its action is the PID parameter selection (Fig. 3). Such a definition of the AGENT1 is able to smoothly vary the PID parameters within the schedule intervals. For the present application AGENT1 is featured by 4 and 5 fuzzy rules for the IGV and VCE controllers, respectively.

The second agent, called AGENT2, has been added to amplify the PID parameter in order to increase the controller promptness. However, such amplification cannot be the same in all the working conditions. In fact, as sketched in Fig. 4, AGENT2 performs a mapping between the absolute value of relative control error and its first time derivative on one side and the amplification factor on the other side. In this way, e.g., if the control error is large and its first time derivative is positive (the control error is increasing), then the amplification has to be large in order reduce the control error. On the other hand, if the control error is large and its first time derivative is negative (the control error is decreasing), then the amplification has to be small in order to avoid undesired oscillating behavior. Therefore, in AGENT2, the perception is represented by the control error and its first time derivative whereas the action consists in the selection of the amplification factor. AGENT2 decisions are based on 48 rules, which cover the range $[0,5]$ for absolute value of relative control error and $[-0.1,0.1]$ for its time derivative. The corresponding multiplier values range in $[1,2.8]$, $[0.02,2]$, and $[1,20]$, for proportional, integral, and derivative actions, respectively.

The third agent, AGENT3, presents the same perception of AGENT2, but it does not affect the PID parameters. On the contrary its action affects directly the control law. In fact, it performs a direct derivative action, adding an extra control effort only when the control error is large and its first time derivative is positive (the control error is increasing). In all the other conditions it presents no action. Its policy function is based on 107 rules, which cover the range $[0.8,1.2]$ for the relative control error and $k_1*[-1,1]$ for the time derivative of its absolute value, where k_1 assumes the values 20 and 100 for the VCE and IGV controller, respectively. The corresponding output values ranges in $k_2*[-0.1,0.1]$, where k_2 assumes the values 10 and 1 for the VCE and IGV controller, respectively.

The last agent, AGENT4, provides the steady state value of the control variable according to the electrical load demand. In other words, AGENT4, carries out an open-loop fuzzy mapping between the external load (disturbance) and the control variable, by means of 13 fuzzy rules. It does not present any feedback validation concerning the control error. Therefore, even if, on the one hand, AGENT4 promptly supplies almost the entire control effort needed in steady state condition for a given load variation, on the other hand, it is unable to generate appropriate control law during dynamic conditions or to annihilate static error. For these reasons AGENT4 is complementary to the first three agents since they act meanly during transient conditions preventing undesired static control errors, whereas the former promptly yields nearly the final value of the control variable.

Two final considerations are in order: first, the MAS structure presents four agents. They do not have the same perception and produce different actions. Moreover, the first two collaborate to improve the readiness and the precision of the existing PID. On the other hand, the others act directly on the control law. Such a MAS does not present a master agent which governs the other agents. On the contrary each agent performs its task independently from each others. Besides, the two MAS structures work independently on the two controlled variables (one can says they constitute a larger MAS) and do not need of any uncoupling procedures that prevent from undesired mutual influence of the two controllers. Such an aspect is very interesting since allows us to design a MAS independently from other MAS structures acting on the same environment.

Mathematical Model of the Gas Turbine

A mathematical model previously developed by the authors (Ref. [18]) numerically simulates the V64.3A gas turbine single shaft gas turbine plant for electric power generation with a 17-stage axial flow compressor. The first four compressor stages are equipped with adjustable guide vanes (IGV) to improve performance at part-load. Seven of the eight turbine rows are cooled, however in this work, for sake of simplicity, the last two stages

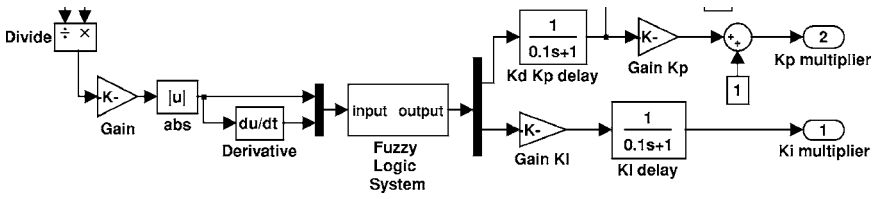


Fig. 4 AGENT2 internal layout

Table 1 Specifications of Ansaldo-Siemens V64.3A Engine

Shaft speed (rpm)	5420
Electrical power (MW)	67.8
Cycle efficiency	35.9
Inlet air flow (kg/s)	189.3
Pressure ratio	16.6
Turbine inlet temperature (K)	1581
Turbine outlet temperature (K)	867

have been considered as adiabatic. The design specifications of the V64.3A engine are summarized in Table 1. The turbine expander has been schematized making use of three different blocks. The first two blocks represent the cooled stage dynamic in which the model proposed by El Mastri (Ref. [19]) for cooled expansion has implemented. The third block simulates the last two adiabatic stages. A model validation is at the moment quite difficult for the lack of experimental data arising from the transient test. Due to this obstacle, a part-load validation has been carried out on the basis of the experimental data published by Jansen et al. in Ref. [20].

Moreover, the temperature transducer dynamic of the shielded thermocouple has been simulated by means of a linear model in s-space characterized by one zero and two poles (Ref. [18]). As far as the variable inlet guide vanes is concerned, the authors simulate its actuator by means of a first order transfer function, limiting both the range extension and the maximum actuation speed.

Dynamic parameters such as temperature transducer transfer function, flame delay, combustion chamber volume, compressor discharge volume, fuel system transfer function, IGV actuator transfer function, shaft and gear inertia are taken from the literature and, when it was possible, from the technical specifications.

Results

The multiagent system control scheme described in the previous sections has been applied to the single-shaft heavy-duty gas turbine in the MIMO mode. In the application, the controlled variables are represented by the shaft rotational speed (power frequency) and stack temperature (related to the overall gas turbine efficiency); on the other hand, the fuel mass flow (VCE) and the variable inlet guide vanes (VIGV) have be chosen as control variables.

In the simulation test, the plant is assumed to undergo to a prescribed time history variation of the electric load (Fig. 5) that, from the control point of view, represents the disturbance variable. As illustrated in Fig. 5, the turbine experiences different types of load variations: in the first part of the test, a full load rejection occurs whereas, in the second part it is subjected to increasing and decreasing linear ramps. The reason for such a prescribed time history variation of the electric load is explained as follows: the aim of the first part of test is not only to prove the capability of the proposed methodology to counterbalance sudden variations of the electrical load, but it is also to cover a wide area of working conditions. On the other hand, the second part of the simulation tests the capability of the MAS control scheme to reject more severe disturbance variations, represented by increasing and decreasing linear ramps. In particular the slope of such ramps has been chosen unusually high in order to provide a more severe test.

A comparison between the proposed methodology, a standard PID controller optimized (in terms of promptness, accuracy, and controlled variable rate of approach to the reference value) for the nominal working condition, and a conventional scheduled PID controller has been presented. Scheduled PID controller shares the same control parameter map with fuzzy AGENT1. Figures 6 and 7 show the results obtained by the application of the MAS control scheme and the PID algorithms. Specifically, Fig. 6 illustrates the control signal fed into the VCE (Figs. 6(*a*) and 6(*b*)) and the normalized turbine rotational speed (Fig. 6(*c*)), whereas Fig. 7 refers to the control signal to the VIGV (Figs. 7(*a*) and 7(*b*)) and to the corresponding normalized stack temperature (measured by means of a temperature transducer dynamic of shielded thermocouple) (Figs. 7(*c*) and 7(*d*)). Such figures report not only the comparison between the MAS control scheme and PID controllers, but they also report each agent contribution to the control variable value.

From Fig. 6(*c*) it is possible to note not only that the MAS control scheme reduce considerably the initial overshoot due to the load rejection (1.6% of the MAS scheme versus 2.5% of the PID schemes), but it also modulates the approach velocity to the reference value preventing a negative overshoot that arises in the standard PID applications. Moreover, even during the increasing and decreasing linear ramps MAS control methodology produces better results with respect to those of the PIDs in terms precision, promptness, and rate of approach to the reference value. A final consideration about the single agent contributions to the control law is in order. The AGENT4 contribution derives from an open-loop fuzzy mapping between the external load (disturbance) and the control variable; for such reason, on one hand, it present a stabilizing effect on the shaft rotational speed, on the other hand, it is unable to promptly react during transients and to annihilate control static errors. On the contrary, the other agents take care of dynamic conditions ensuring a zero static error. In particular AGENT1 and AGENT2 collaborate tuning the PID components, whereas the AGENT3 contribution, due to its derivative nature, modulates the approach velocity of the controlled variable to its reference value.

As far as the stack temperature is concerned, Fig. 7 reports not

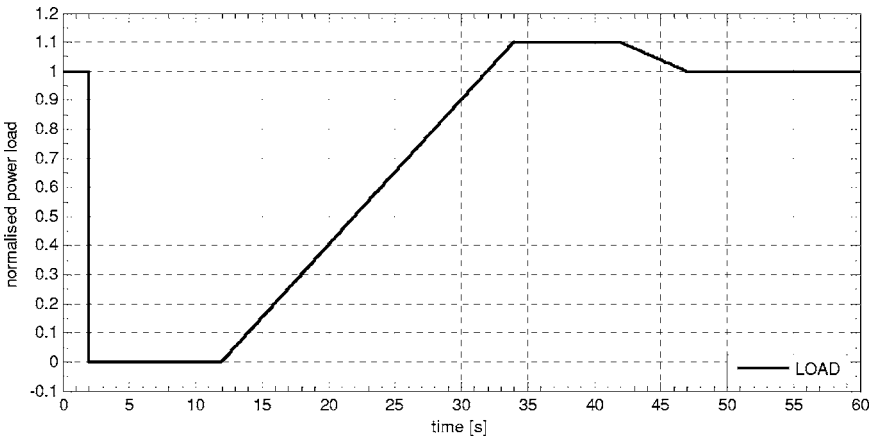


Fig. 5 Gas turbine load history

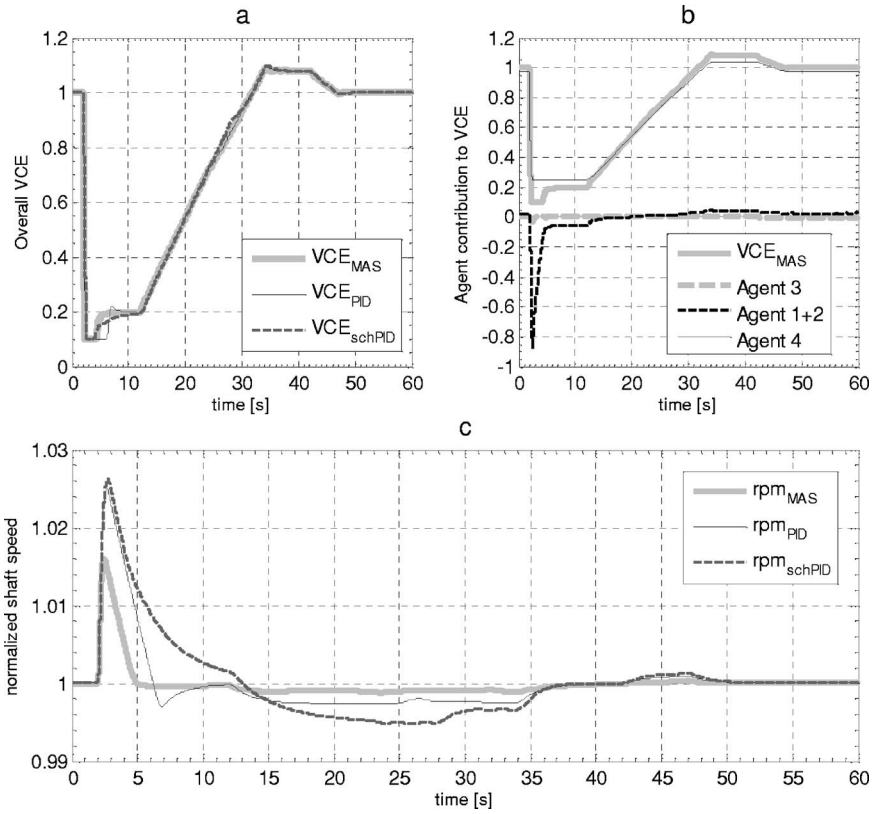


Fig. 6 Comparison MAS versus PID approaches: shaft rotational speed results

only the VIGV signal (Figs. 7(*a*) and 7(*b*)), the corresponding normalized stack temperature value, measured by means of a temperature transducer dynamic of shielded thermocouple (Fig. 7(*c*)), but it also shows, for completeness, the actual stack temperature evaluated by the gas turbine model for the MAS and PID approaches (Fig. 7(*d*)). In the first part of the test it possible to notice only minor differences between the performance of MAS and PID controllers. This is mainly due to the limitation of the maximum actuation speed of the variable inlet guide vanes. In fact, the load rejection causes all the controllers to rapidly close the inlet guide vanes, reaching the maximum allowed velocity. Nevertheless, a slightly different behavior arises between the MAS and the PID controllers. Such a difference can be ascribed, on the one hand, on the different behavior of the rotational speed regulated by the MAS and PID approach; on the other hand, especially focusing on the actual stack temperature, it is possible to point out a better response of the MAS methodology in terms of promptness and accuracy. Right after the load rejection, all the controllers incur in a second inlet guide vanes limitation regarding the range extension. In fact, the VIGV control system (both MAS and PIDs) has to keep the exit temperature constant from 50% to 110% of the nominal load; outside of this range the IGV position is kept constant. The final part of the test clearly demonstrates the better capabilities of the MAS system with respect the PID ones. In fact, the former presents neither overshoot nor oscillating behavior, whereas the scheduled PID shows a 7.9% overshoot and the standard PID is affected by a small oscillating behavior. In spite of the small oscillations produced by the standard PID controller in terms of measured stack temperature, its real value experiences a larger temperature variation which corresponds to a more severe thermal stress. Moreover, the MAS controller is also characterized by a faster approach to the reference value.

A further consideration is in order: from (Figs. 6(*b*) and 7(*b*)), it is worthy to note that not all the agents of the MAS control scheme work in all the gas turbine working conditions. In particular, AGENT2 and AGENT3 are disabled whenever the first time derivative of the control error assumes small values. In such conditions only AGENT1 and AGENT4 are enabled and the MAS scheme behaves as a fuzzy system that provides smooth scheduled PID control parameters in which the steady state value of the control variable is provided by AGENT4. Therefore in the quasistationary condition the MAS is equivalent to a scheduled PID with the same characteristic of precision, robustness, and stability.

In addition, Fig. 8 shows what happens when one or more agents are disabled. Specifically Fig. 8 reports the results, in terms of shaft rotational speed, obtained by disabling, in the VCE controller, only AGENT2, both AGENT2 and AGENT4, and, finally, AGENT2, AGENT3, and AGENT4. Such results, on the one hand, assess the decentralized structure of the MAS controller; even if by disabling agents the overall performance decreases, the main tasks of the controller (i.e., control stability) is ensured. Moreover, the modular structure is reflected by the collaborative action of each agent which always contributes to accuracy improvements, independently from each other. Consequently, the design of the MAS controller is carried out by a modular approach since the single agent does not have to face the whole control problem, but it has to accomplish a well defined task. The good level of robustness and accuracy is confirmed by Fig. 9 which presents the stack temperature time history obtained by disabling the VCE controller agents. Specifically, Fig. 9(*a*) refers to the stack temperature measured by means of the shielded thermocouple, whereas Fig. 9(*b*) illustrates the actual stack temperature evaluated by the gas turbine model. In these conditions, the VIGV MAS controller is still featured by all agents and therefore it is able to counterbalance the effects coming from VCE agents disabling. In fact, only minor differences may be observed between the diagrams reported in Fig. 9, although the best performance is still to ascribe to the full MAS VCE controller.

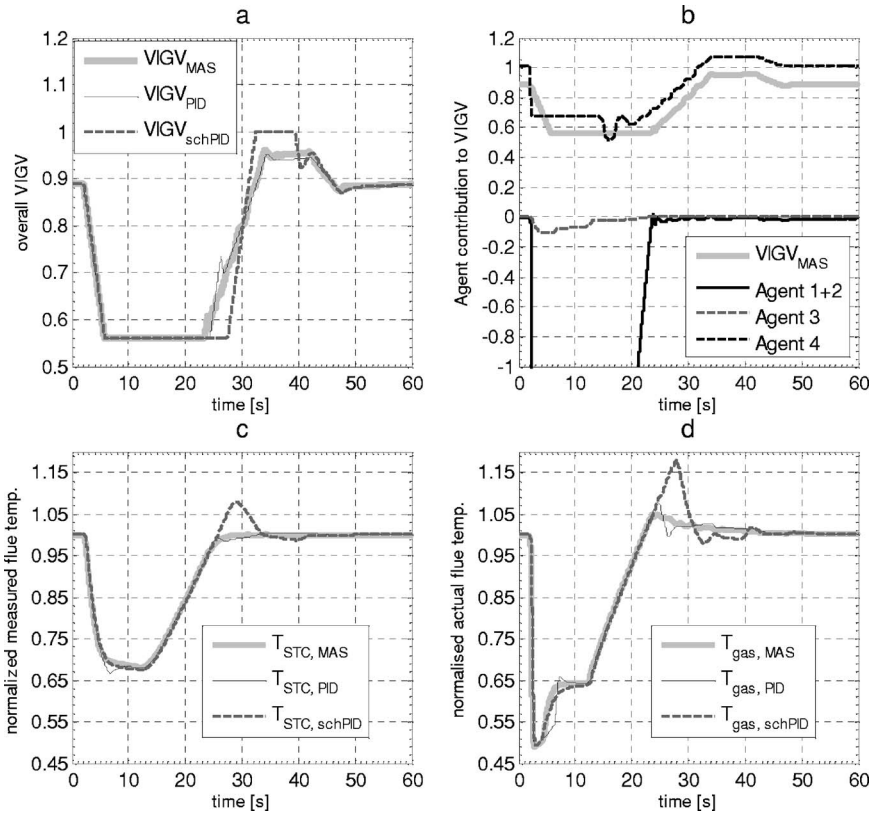


Fig. 7 Comparison MAS versus PID approaches: stack temperature results

Conclusions

In the present work, a MAS approach has been proposed to manage the control process of a single-shaft heavy-duty gas turbine. MAS control scheme has been applied in multi input multi output mode, to control the the fuel mass flow and the angle of variable inlet guide vanes the shaft rotational speed (power frequency) and stack temperature (related to the overall gas turbine efficiency).

It has been shown that the MAS controller performs better than a traditional PID controller in terms of promptness but it also shows a good level of adaptivity, obtained by using agents operating as fuzzy logic systems. Thanks to this choice the resulting MAS has shown a high degree of flexibility. Moreover the control task was decomposed in simpler and smaller subtasks which were fulfilled by single agents; moreover it was possible to blend computed rules with heuristic ones. The proposed control also appeared robust in the sense that it showed its capability to work well also when one or more agents are disabled or do not work properly.

Moreover, the multiagent system presents further attractive advantages with respect the single-agent approach: it reduces the computational effort due to the asynchronous and parallel computation. In addition the multiagent system allows us to increase scalability and flexibility, since it is easy to add new agents to the system, maintaining a simple architecture for each agent.

Acknowledgment

Part of the present work has been carried out within the research program "*Dynamic Modelling of Energy Systems*" funded

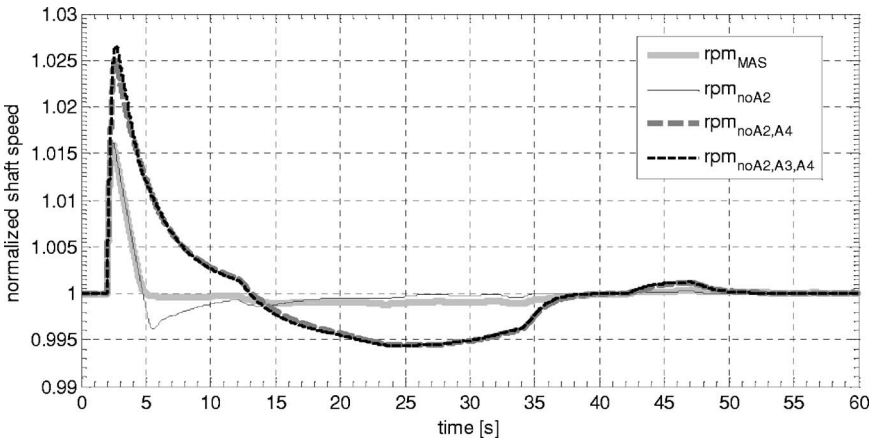


Fig. 8 Full MAS versus disabled agents on VCE: shaft rotational speed

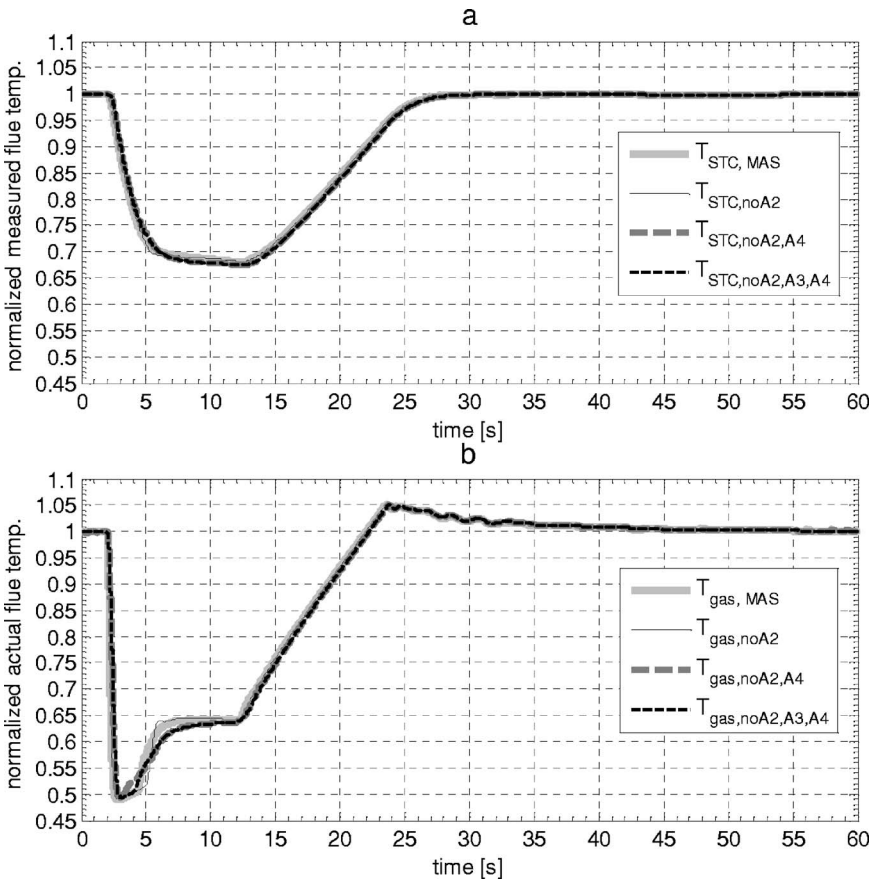


Fig. 9 Full MAS versus disabled agents on VCE: stack temperature

by The Italian Ministry for Education, University and Research (MIUR).

Nomenclature

a = agent action
k = iterative index variable
o = agent perception
$\mathbf{u}$ = fuzzy logic output vector
$\mathbf{z}$ = fuzzy logic input vector
$\hat{z}$ = mean value of the input fuzzy Gaussian membership function
$\tilde{\mathbf{A}}$ = antecedent fuzzy sets
$\hat{A}$ = mean value of the antecedent fuzzy Gaussian membership function
N = number of fuzzy logic rules
R = fuzzy logic base of rules
β = consequent fuzzy sets
μ = fuzzy logic membership function
π = agent policy function
σ = variance of fuzzy Gaussian membership function
DARMA = deterministic auto regressive moving average
FLS = fuzzy logic system
MAS = multiagent system
PID = proportional integrative derivative controller
MIMO = multi input multi output
VCE = fuel valve actuator
VIGV = variable inlet guide vanes

References

[1] Daniele, C. J., and Krosel, S. M., 1979, "Generation of Linear Dynamic Models From a Digital Nonlinear Simulation," NASA Technical Paper No. 1388, NASA.
[2] Smith, D. L., and Stammetti, V. A., 1990, "Sequential Linearization as an Approach to Real-Time Marine Gas Turbine Simulation," J. Eng. Gas Turbines Power, **112**, pp. 187–191.
[3] Dambrosio, L., Camporeale, S. M., and Fortunato, B., 2000, "Performance of Gas Turbine Power Plants Controlled by One Step Ahead Adaptive Technique," ASME Paper No. GT-2000-037.
[4] Dambrosio, L., and Fortunato, B., 1999, "One-Step-Ahead Adaptive Control of a Wind-Driven, Synchronous Generator System," Energy, **24**, pp. 9–20.
[5] Goodwin, G. C., and Sin, K. S., 1984, *Adaptive Filtering Prediction and Control*, Prentice-Hall, Englewood Cliffs, NJ.
[6] Dambrosio, L., Camporeale, S. M., and Fortunato, B., 2002, "Control of Gas Turbine Power Plants by Means of the Weighted One Step Ahead Adaptive Technique," Proc. Inst. Mech. Eng. Part I–Journal of System and Control Engineering, **216**, pp. 317–324.
[7] Zawia, S. M., and Al-Khodari, M. S. B., 2004, "Dynamic Modeling and Self-Tuning Control of a Gas Turbine," ASME Paper No. GT-2004-53530.
[8] Mu, J., and Rees, D., 2004, "Nonlinear Model Predictive Control for Gas Turbine Engines," ASME Paper No. GT-2004-53146.
[9] Brunell, B. J., Viassolo, D. E., and Prasanth, R., 2004, "Adaptation and Nonlinear Model Predictive Control of an Aircraft," ASME Paper No. GT-2004-53780.
[10] Dambrosio, L., Pascazio, G., and Fortunato, B., 2005, "Fuzzy Logic Controller Applied to Variable Geometry Turbine Turbocharger," Proc. Inst. Mech. Eng., Part D (J. Automob. Eng.), **219**, pp. 1347–1360.
[11] Martucci, A., and Volponi, A. J., 2000, "Fuzzy Fuel Flow Selection Logic for a Real Time Embedded Full Authority Digital Engine Control," ASME Paper No. 2000-GT-0046.
[12] Mu, J., and Rees, D., 2003, "Optimum Gain-Scheduling PID Controllers for Gas Turbine Engines Based on Narmax and Neural Network Models," ASME Paper No. GT-2003-38667.
[13] Vroemen, B. G., van Essen, H. A., van-Steenhoven, A. A., and Kok, J. J., 1999, "Nonlinear Model Predictive Control of a Laboratory Gas Turbine Installation," J. Eng. Gas Turbines Power, **121**, pp. 629–634.
[14] Martucci, A., Fuller, J., Dorobantu, E., and Rahnamai, K., 2004, "The Effect of Terminal Weight on the Prediction Horizon of a Gas Turbine Engine Using Model Predictive Control," ASME Paper No. GT2004-53009.

[15] Vlassis, N., 2003, "A Concise Introduction to Multiagent Systems and Distributed," *AI Intelligent Autonomous Systems Informatics Institute*, University of Amsterdam.
[16] Russell, S. J., and Norvig, P., 2003, *Artificial Intelligence: A Modern Approach*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ.
[17] Mendel, J. M., 1995, "Fuzzy Logic Systems for Engineering: A Tutorial," Proc. IEEE, **83**, pp. 345–377.
[18] Camporeale, S. M., Fortunato, B., and Mastrovito, M., 2002, "A High Fidelity Real Time Simulation Code of Gas Turbine Dynamics for Control Applications," ASME Paper No. GT-2002-30039.
[19] El Mastri, M. A., 1988, "On the Thermodynamics of Gas Turbines Cycles—Part II: A Model of Expansion in Cooled Turbines," J. Eng. Gas Turbines Power, **110**, pp. 201–209.
[20] Jansen, M., Schulenberg, T., and Waldinger, D., 1992, "Shop Test Results for the v64.3 Gas Turbine," J. Eng. Gas Turbines Power, **114**, pp. 676–681.

Hybrid Kalman Filter Approach for Aircraft Engine In-Flight Diagnostics: Sensor Fault Detection Case

Takahisa Kobayashi
ASRC Aerospace Corporation,
21000 Brookpark Road,
Cleveland, OH 44135

Donald L. Simon
U.S. Army Research Laboratory,
Glenn Research Center,
21000 Brookpark Road,
Cleveland, OH 44135

In this paper, a diagnostic system based on a uniquely structured Kalman filter is developed for its application to in-flight fault detection of aircraft engine sensors. The Kalman filter is a hybrid of a nonlinear on-board engine model (OBEM) and piecewise linear models. The utilization of the nonlinear OBEM allows the reference health baseline of the diagnostic system to be updated, through a relatively simple process, to the health condition of degraded engines. Through this health baseline update, the diagnostic effectiveness of the in-flight sensor fault detection system is maintained as the health of the engine degrades over time. The performance of the sensor fault detection system is evaluated in a simulation environment at several operating conditions during the cruise phase of flight. [DOI: 10.1115/1.2718572]

Keywords: in-flight fault detection, on-board engine model, Kalman filter, flight safety

Introduction

In-flight diagnostics of aircraft gas turbine engines is a critical task for the improvement of aviation safety. The capability to detect and/or isolate any faults, which may cause the engine to operate at an undesirable condition during flight, can improve not only the safety but also the efficiency of engine operation. Since the diagnostic system results can influence the follow-up actions taken by the flight crew or control system, it is critical that they be highly reliable. In-flight diagnostic systems, therefore, must be designed with robustness to nonfault-related factors which exist in the real environment and can potentially mislead diagnostic systems to generate incorrect results.

In-flight diagnostic systems are, in general, designed at a nominal health, or nondegraded, condition. This design condition becomes a reference health baseline for the diagnostics; any observed deviation in engine outputs from their reference condition values indicates the presence of a fault. In-flight diagnostic systems can perform effectively as long as the health of a real engine remains in the vicinity of the reference health baseline, thereby making engine output deviations prominent when a fault takes place.

As the real engine degrades over time, in-flight diagnostic systems may lose their effectiveness. Engine health degradation is a normal aging process that occurs to all aircraft engines due to usage, and therefore is not considered as a fault. However, similar to various faults, degradation causes the engine outputs to deviate from their reference condition values. When engine output deviations eventually exceed a certain level, the diagnostic systems may misinterpret the health degradation as a fault and consequently generate a false alarm.

One approach to maintaining the effectiveness of in-flight diagnostics applied to degraded engines is to periodically update or re-design the diagnostic algorithms based on the estimated amount of health degradation. Health degradation can be estimated by trend monitoring systems using postflight data [1–3]. Through the update based on the estimated health degradation, the health baseline of an in-flight diagnostic system can be shifted to the vicinity of the degraded engine, and therefore the system is able to effectively detect the presence of a fault. One issue with this approach is its practicality. Depending on the complexity of the diagnostic algorithms, the update process may take too much time and thus may be impractical.

To address the above-mentioned issue, a Kalman filter based in-flight sensor fault detection system is developed in this paper. The proposed Kalman filter is composed of a nonlinear on-board engine model (OBEM) and piecewise linear state-space models, which include Kalman gain matrices. With this architecture, the system update to account for engine health degradation is achieved through a relatively simple process: by feeding the estimated health degradation values into the OBEM. Thus, fault detection can be accomplished even without updating linear models and associated Kalman gains as the real engine degrades over time.

In the following sections of this paper, the problem setup and the design approach for an in-flight sensor fault detection system are described, followed by the application of the design methodology to a large commercial aircraft engine model. To validate its diagnostic effectiveness, the in-flight sensor fault detection system is evaluated in a simulation environment using various fault and degradation scenarios at multiple operating conditions during the cruise phase of flight.

Problem Setup for an In-Flight Fault Detection System

The objective of an in-flight fault detection system is to detect faults as quickly as possible from the observed engine outputs while avoiding false alarms and missed detections. Since false alarms are generated as the result of misinterpretation of nonfault-related factors, it is important to understand the influence of such factors on engine outputs. Engine health degradation is one of the nonfault-related factors that can cause false alarms. As shown in Fig. 1, engine health degradation is described as gradual deviations of health parameters from the initial healthy baseline. Health parameters are efficiencies and flow capacities of engine components such as compressors and turbines, and they indicate the health of such components. As they deviate from the initial healthy baseline, engine outputs will also deviate from their nomi-

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received November 16, 2006; final manuscript received November 17, 2006. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2006: Land, Sea and Air (GT2006), May 8–11, 2006, Barcelona, Spain. Paper No. GT2006-90870.

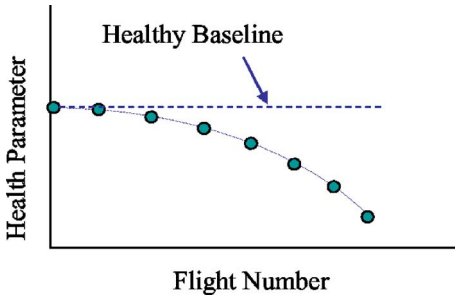


Fig. 1 Engine health degradation

nal condition values. Since degradation is a normal aging process that all aircraft engines will experience due to usage, it is not considered as a fault, whereas a fault is an abnormal and unexpected event. However, as the engine output deviations increase with time due to the progression of health degradation, it becomes difficult to distinguish the presence of faults from health degradation through the observation of engine outputs. As a result, an in-flight fault detection system loses its diagnostic effectiveness as the engine degrades over time. An example of such loss of diagnostic effectiveness, due to health degradation, is shown in Refs. [4,5] for the case of sensor fault diagnostics based on neural networks.

To reduce the influence of health degradation on in-flight diagnostic performance, the fault detection system needs to be updated periodically based on the estimated amount of health degradation, as shown in Fig. 2. Health degradation can be estimated by trend monitoring systems using postflight data [1–3]. Since health degradation progresses gradually with time, it is expected that the actual change in health parameters is small for a number of flights. Therefore, the process of estimating health degradation and updating the in-flight diagnostic system needs to be accomplished once per a number of flights. When the updating process is complete, the estimated health condition becomes the new reference baseline for the fault detection system, as shown in Fig. 3, until the next update process is completed. Through this periodic baseline update, the in-flight diagnostic system can operate in the vicinity of the degraded engine and thus maintain its effectiveness for fault detection.

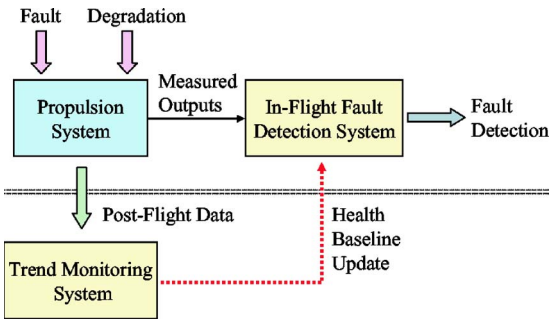


Fig. 2 Process of health baseline update

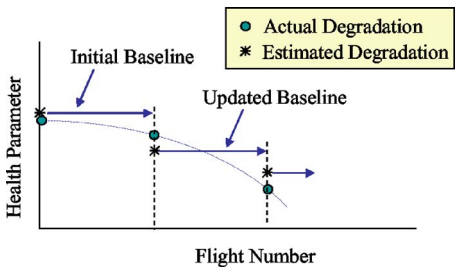


Fig. 3 Baseline update using estimated health degradation

In this paper, it is assumed that a trend monitoring system, which is capable of estimating engine health degradation (health parameters), is available. Moreover, it is assumed that this trend monitoring system will achieve a certain level of precision in its estimation performance. The rest of the paper focuses on the development of the in-flight sensor fault detection system based on the hybrid Kalman filter approach.

Development of an In-Flight Sensor Fault Detection System

The in-flight fault detection system described in the following is based on the Kalman filter approach with a unique hybrid architecture. The Kalman filter is composed of a nonlinear OBEM and piecewise linear state-space models, which include Kalman gain matrices. The OBEM is a physics-based model designed to run in real time, while the piecewise linear state-space models are derived off-line from the OBEM at the nominal health baseline. These two main components are merged together to form the "hybrid" Kalman filter. Based on the residuals generated by the hybrid Kalman filter, a fault indicator signal is constructed for detecting faults.

Hybrid Kalman Filter Design. The design steps for the hybrid Kalman filter are exactly the same as those for the general linear Kalman filter. First, a nonlinear plant model is linearized at operating points. Then, Kalman gains are computed based on the linear representations of the plant model. When implemented, however, linear models and associated Kalman gains are integrated with the nonlinear plant model. An aircraft engine under consideration is represented by a nonlinear model of the following form:

$$\dot{x} = f(x,h,u_{\text{cmd}},e)$$

$$y = g(x,h,u_{\text{cmd}},e) + v \tag{1}$$

where x, h, u_{cmd}, and e represent the vectors of state variables, health parameters, control command inputs, and environmental parameters, respectively. For given input values, the nonlinear functions f and g generate the vectors of state derivatives $\dot{x}$ and sensor outputs y. The sensor outputs are corrupted by the white noise vector v. By linearizing the engine model at a reference health baseline (e.g., nominal health condition) and also at a specific environmental condition, the following state-space equations are obtained:

$$\dot{x} = A(x - x_{ss}) + B(u_{\text{cmd}} - u_{ss}) + L(h - h_{\text{ref}})$$

$$y - y_{ss} = C(x - x_{ss}) + D(u_{\text{cmd}} - u_{ss}) + M(h - h_{\text{ref}}) + v \tag{2}$$

where A, B, C, D, L, and M are the state-space matrices with appropriate dimensions. The vectors x_{ss}, y_{ss}, u_{ss} contain the steady-state values at which the engine model is trimmed for linearization. The vector h_{ref} represents a reference health baseline. The Kalman gain is computed based on the matrix pair $[A, C]$, and the linear Kalman filter equation is given as follows:

$$\dot{\hat{x}} = A(\hat{x} - x_{ss}) + B(u_{\text{cmd}} - u_{ss}) + K(y - \hat{y})$$

$$\hat{y} - y_{ss} = C(\hat{x} - x_{ss}) + D(u_{\text{cmd}} - u_{ss}) \tag{3}$$

The vectors $\hat{x}$ and $\hat{y}$ represent the estimates of the state variables and sensor outputs, respectively. The matrix K represents the Kalman gain. In order for the Kalman gain to converge, the matrix pair $[A, C]$ must be observable.

It should be noted that the linear Kalman filter in Eq. (3) does not account for the influence of health parameter deviations from the reference health baseline in Eq. (2). Since the Kalman filter is designed with some robustness to system uncertainty in the form

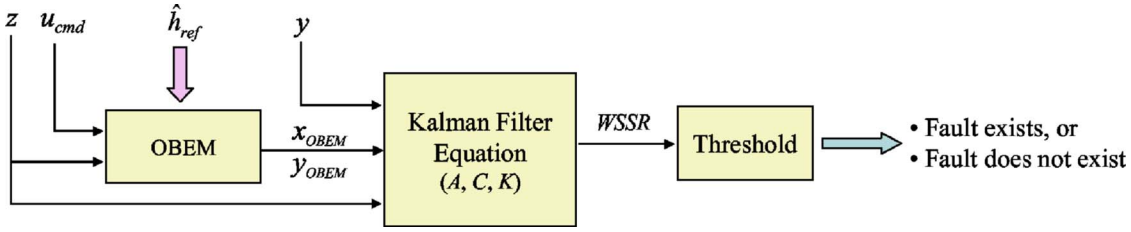


Fig. 4 Overall architecture of the in-flight fault detection system

of process noise, it has robustness to health parameter deviations to some extent. However, the Kalman filter does not have the level of robustness that can handle the full health deterioration that an engine will experience over its lifetime. Therefore, as discussed in Refs. [6,7], the Kalman filter must be updated periodically based on the estimated health condition in order to allow the Kalman filter to operate in the vicinity of a real engine as it degrades. The process of health baseline update for the general linear Kalman filter is described in the following four steps: (1) estimate the health degradation, (2) trim the closed-loop engine model at the new reference health baseline (estimated health condition), which should be close to the actual health condition, and generate the steady-state vectors, (3) linearize the open-loop engine model and generate state-space matrices, and then (4) compute the Kalman gain. Step 1 can be done off-line by a trend monitoring system which monitors the engine health degradation over time. Steps 3 and 4 may not be necessary according to Ref. [6], but step 2 alone can be a time-consuming, troublesome process, especially when many operating points must be covered over the flight envelope. Thus, it is desirable to simplify the update process in order to make it feasible in the real application environment.

Insight into simplifying the health baseline update process can be found from past studies. Reference [6] indicated that the performance of the linear Kalman filter when applied to degraded engines can be improved significantly just by updating the steady-state vectors (x_{ss}, y_{ss}, u_{ss}) to the new values derived at degraded conditions. Moreover, Ref. [8] demonstrated that the constant gain extended Kalman filter can operate over a wide operating range despite its simple architecture, which basically combines a nonlinear engine model with a single Kalman gain matrix computed at a single operating point. These studies indicate that the Kalman gain itself is not of primary importance to operate the Kalman filter in the environment where various elements, such as health or flight condition, are changing. Rather, the accuracy of the plant model is of primary importance.

Based on the above-mentioned knowledge, the hybrid Kalman filter is developed by replacing the steady-state vectors of Eq. (3) with the following nonlinear OBEM:

$$\dot{x}_{\mathrm{OBEM}} = f(x_{\mathrm{OBEM}}, \hat{h}_{\mathrm{ref}}, u_{\mathrm{cmd}}, z)$$

$$y_{\mathrm{OBEM}} = g(x_{\mathrm{OBEM}}, \hat{h}_{\mathrm{ref}}, u_{\mathrm{cmd}}, z) \tag{4}$$

where the vector $\hat{h}_{\mathrm{ref}}$ represents the health condition estimated by a trend monitoring system, which is updated once per a number of flights. The vector z represents the measured parameters which define the flight condition. By integrating the OBEM and linear state-space models, the following hybrid Kalman filter is formed:

$$\dot{\hat{x}} = A(\hat{x} - x_{\mathrm{OBEM}}) + K(y - \hat{y})$$

$$\hat{y} = C(\hat{x} - x_{\mathrm{OBEM}}) + y_{\mathrm{OBEM}} \tag{5}$$

In Eq. (5), the steady-state vectors which appeared in Eq. (3) were replaced by the state variables and engine outputs generated by the OBEM. Furthermore, the control command inputs and associated matrices B and D in Eq. (3) do not appear in Eq. (5) since the effect of control command inputs is accounted for by the OBEM as seen in Eq. (4).

There are a few things that should be noted about the hybrid structure. First, the hybrid Kalman filter depends on the OBEM but not vice versa. The OBEM runs in parallel with the actual engine at the estimated health condition without receiving any feedback signals from the hybrid Kalman filter. Therefore, the numerical stability of the OBEM is not influenced by the performance of the hybrid Kalman filter. The objective of the OBEM is to generate the state variables and sensor outputs at the estimated health condition. By updating the health condition of the OBEM, its state variables and sensor outputs can be brought close to the values of the degraded engine. Since health condition mismatches still exist between the OBEM and the degraded engine due to estimation errors, sensor output mismatches also exist between them. The objective of the hybrid Kalman filter, or specifically its linear component, is to further improve the sensor output matching between its estimates and the measured values through the tuning of the state variable estimates. As long as the OBEM operates in the vicinity of the degraded engine (i.e., health condition estimation errors are small), the hybrid Kalman filter will maintain its accurate sensor output estimation performance.

Construction of Fault Indicator Signal. The validation of the Kalman filter estimates is generally done by checking residuals, or the differences between the measured and estimated sensor output values. If residuals are large, it can be considered that the Kalman filter is generating inaccurate sensor output estimates because of the presence of an anomaly, such as a sensor fault, that was not accounted for in the Kalman filter design. To indicate the presence of a fault, a weighted sum of squared residuals (WSSR) is computed as follows:

$$\mathrm{WSSR} = (y - \hat{y})^T \Sigma^{-1} (y - \hat{y}) \tag{6}$$

where

$$\Sigma = \mathrm{diag}[\sigma^2]$$

The vector σ represents the standard deviation of the sensor measurements. The square matrix Σ normalizes the residual vector $(y - \hat{y})$. Since the hybrid Kalman filter design discussed in the previous section does not account for the presence of a sensor fault, the value of the fault indicator signal, WSSR, should increase when such a fault occurs in the system. The next step for detecting a fault is to compare the fault indicator signal to a preestablished detection threshold. When the fault indicator signal exceeds the detection threshold, it is considered that a fault indeed exists in the system. The establishment of a detection threshold can be based on statistics or achieved through systematic analysis.

Overall Architecture of the In-Flight Sensor Fault Detection System. The overall architecture of the in-flight sensor fault detection system is shown in Fig. 4. The fault detection system receives the measured variables (y and z) and control commands (u_{cmd}) from the on-board digital engine control unit, while receiving a periodically updated health baseline ($\hat{h}_{\mathrm{ref}}$) from the trend monitoring system. The output of the fault detection system is one of two possibilities: fault exists or fault does not exist. The fault detection system reveals only the existence of a fault, not its iden-

Table 1 Engine model variables

State variables	XNL, XNH, TMHS23, TMHS3 TMHSBL, TMHSBC, TMHS41 TMHS42, TMHS5
Health parameters	FAN efficiency, FAN flow capacity BST efficiency, BST flow capacity HPC efficiency, HPC flow capacity HPT efficiency, HPT flow capacity LPT efficiency, LPT flow capacity
Actuators	WF36, VBV, VSV
Environmental parameters	Altitude, Mach number Ambient temperature

Table 2 Standard deviations of controls and diagnostics sensors (σ in percent of steady-state values at ground maximum power condition)

Sensors (y)	σ (%)
XN12	0.25
XN25	0.25
P25	0.50
T25	0.75
PS3	0.50
T3	0.75
T49	0.75

tity or severity.

Since the current system is designed without any restrictions on the types of faults it should detect, it may detect faults in something other than sensors. Additional faults that an engine may encounter during flight include actuator and component faults. Similar to the case of sensor faults, actuator and component faults must be detected through the observation of sensor output deviations. However, the level of sensor output deviations due to these faults heavily depends on location of sensors. Some of the actuator and component faults, therefore, may be very difficult to detect with a given set of sensors unless their fault magnitude is significantly large. Although it is desirable to detect any type of fault, the focus of the current system is placed on sensor fault detection.

Fault detection is the first step in the diagnostic process. It only reveals the existence of a fault without identifying the nature of the detected fault. After fault detection comes fault isolation [6,7] and then fault accommodation [9,10].

Application of the Hybrid Kalman Filter Methodology to Aircraft Engine Model

In the following, the in-flight sensor fault detection system based on the hybrid Kalman filter is applied to an aircraft engine model. A description of the engine model is given first, followed by a discussion of the piecewise linear models. Then, the derivation of the fault detection threshold is discussed.

Engine Model. The engine model used in this paper is a nonlinear simulation of an advanced high-bypass turbofan engine, a typical power plant for a large commercial aircraft. This engine model has been constructed as a component level model (CLM), which consists of the major components of an aircraft engine. The CLM represents highly complex engine physics while designed to run in real time. Engine performance deviations from the nominal health baseline are modeled by adjustments to efficiency and flow capacity scalars of the following five components: fan (FAN), booster (BST), high-pressure compressor (HPC), high-pressure turbine (HPT), and low-pressure turbine (LPT). There are a total of 10 of these adjustments which are called health parameters. The engine state variables, health parameters, actuator variables, and environmental parameters are shown in Table 1.

There are a total of 11 measured parameters (y and z in Fig. 4) that are available to the digital engine control unit of this engine. Table 2 shows seven sensors (y) along with their standard deviations given in percent of steady-state values at the ground maximum power condition. The control actions and diagnostics are based on those sensed variables. Table 3 shows four additional measured parameters (z) along with their standard deviations given in their actual units. These four parameters indicate the ambient and engine inlet conditions. Some of them (T_{amb}, P_{amb}, T2) are equivalent to the environmental parameters in Table 1, while T2 and P2 are used for parameter corrections [11].

This nonlinear engine model is used to represent both the actual engine and the OBEM in the subsequent sections. Typically, there will be a mismatch between the model and the actual engine due to modeling errors and unmodeled elements. The influence of such a mismatch will be assessed in the later section. The engine model representing an actual engine operates at given health conditions, and its flight condition is specified by the three environmental parameters in Table 1. The OBEM operates at estimated health conditions, and its flight condition is specified by three measured parameters: T_{amb}, P_{amb}, and T2. The actual engine and the OBEM receive the same three control commands (Table 1). In the current control architecture, the power lever angle (PLA) is converted to desired corrected fan speed (an indicator of thrust). The control system adjusts three actuation values to cause the corrected measured fan speed to match the desired value. The closed-loop engine simulation runs with a 0.02 s time step.

Piecewise Linear Model Design. The linear component of the hybrid Kalman filter (A, C, K in Eq. (5)) is designed using the nonlinear engine model through the following steps. The nonlinear engine model is first linearized at specific operating conditions. For each of the linear engine models, a Kalman gain is computed. Then, the piecewise linear models are saved in table lookup form. As the operating condition moves from one point to another, the piecewise linear models are interpolated based on a scheduling parameter as discussed in Ref. [7].

The unique aspect of the hybrid Kalman filter design is that the piecewise linear models are integrated with the OBEM, instead of the steady-state vectors for the case of a pure piecewise linear Kalman filter design [7,12]. As discussed earlier, having an accurate plant representation is of primary importance in the Kalman filter operation. Because of the usage of the OBEM, which is a good representation of an actual engine, the number of operating conditions at which the piecewise linear models are derived does not have to be as many as the case for a pure piecewise linear Kalman filter design. In this paper, the piecewise linear models are generated along the steady-state power setting line at a cruise condition. For the interpolation of the piecewise linear models, the corrected fan speed (XN12) is used as the scheduling parameter. A preliminary study indicated that this specific hybrid Kalman filter design maintains its accurate estimation performance over a wide operating range. When a similar study was done for the hybrid Kalman filter using piecewise linear models generated at sea level static condition, the estimation performance was not as good as the case of cruise condition design. Better estimation performance may be achieved by linking linear models generated at multiple flight conditions, such as climb and cruise. However, it is desirable to keep the number of piecewise linear models as small as possible, since a lesser number of linear models reduces design

Table 3 Standard deviations of ambient and engine inlet sensors (σ in actual units)

Sensors (z)	σ
T_{amb}	5.0°F
P_{amb}	0.1 psi
T2	5.0°F
P2	0.1 psi

complexity and also improves execution speed.

When the hybrid Kalman filter was implemented in a simulation environment, the system was discretized to run at the frequency of 50 Hz. The parameters used by the hybrid Kalman filter algorithm are corrected based on the engine inlet condition T2 and P2.

Selection of the Detection Threshold. The selection of the detection threshold is a critical part in the diagnostic system design. Setting the threshold at a low value increases the chance of detecting faults but also increases the chance of generating false alarms. Conversely, setting the threshold at a high value decreases the chance of generating false alarms but also decreases the chance of detecting faults. As such, the balance between true positive (fault detection) and false positive (false alarms) is adjusted by the threshold. It is statistically impossible to achieve zero false alarm rates while detecting any faults, but it is the general consensus that the false alarm rate should be maintained as low as possible. Keeping that in mind, the threshold is determined in this section.

A false alarm is the result of misinterpretation of nonfault-related factors[1] which exist in various forms to various degrees. If the influence of such factors on the fault indicator signal is known, a threshold can be derived from that knowledge. Among the numerous nonfault-related factors that can cause false alarms, health condition mismatches are used as an example to derive the threshold.

As discussed earlier, in the current approach, the health baseline of the OBEM must be updated periodically as the health of the real engine degrades gradually with time. However, the health baseline update will never be exact in the real environment, and therefore, health condition mismatches will always exist between the OBEM and the real engine. These mismatches can cause the fault indicator signal (WSSR in Eq. (6)) to increase. If the threshold is set to a point higher than the maximum value that the fault indicator signal can reach due to health condition mismatches, then false alarms can be avoided at least for the case of health condition mismatches.

To investigate the influence of health condition mismatches on the fault indicator signal, 300 cases of health degradation and associated health condition estimates were first generated. Three hundred cases of health degradation were created by randomly shifting all 10 health parameters shown in Table 1. The deviation values were uniformly distributed over the range from 1% to 5%, and this level of deviation is beyond the typical level of engine-to-engine variation due to manufacturing tolerance. Estimated health conditions were created by adding estimation errors to the 300 cases of health degradation. The estimation error for each health parameter was a random number with uniform distribution over the range of ±0.5%. As mentioned earlier, it is assumed that a trend monitoring system is available, and its estimation accuracy is assumed to be within ±0.5% from the actual values.

Using the above-mentioned 300 health condition mismatch cases, the engine and the hybrid Kalman filter were run for 100 s at a specific operating point. For each mismatch case, the hybrid Kalman filter generated a time history of the fault indicator signal. Then, the maximum value that the fault indicator signal reached during the 100 s run was saved for each of the 300 cases. Based on the maximum WSSR values for the 300 cases, a histogram was generated to investigate the variation of the fault indicator signal due to health condition mismatches. This process was then repeated at various flight conditions with various power settings. Examples of the histograms generated at three power settings (PLA=65, 69, 70) at a cruise condition are shown in Fig. 5. In the current implementation, the fault indicator signal (WSSR) has been scaled and also processed by a low-pass filter with a cutoff frequency of 0.1 rad/s. The arrows in Fig. 5 indicate the largest maximum WSSR value among the 300 cases.

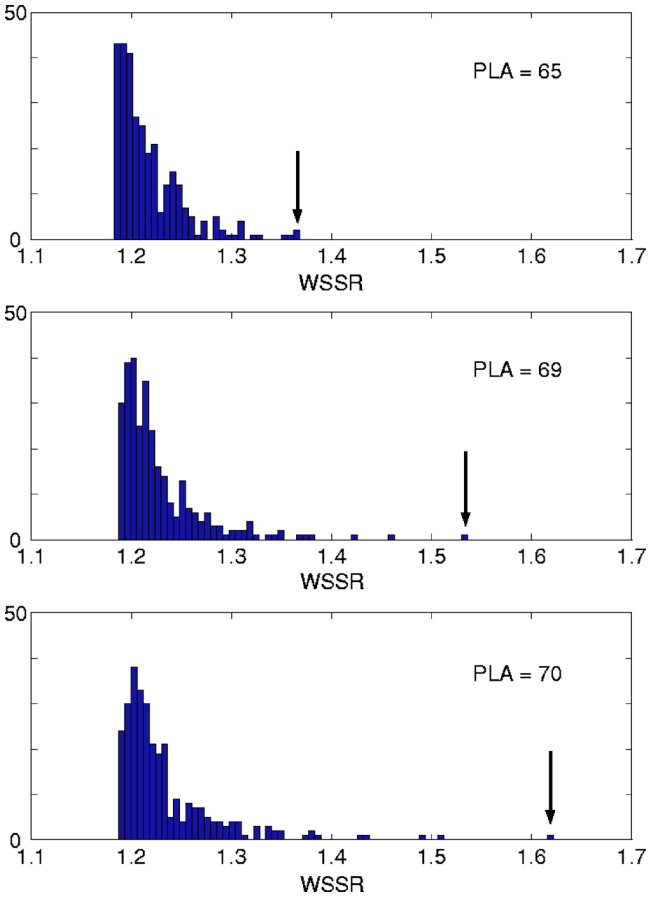


Fig. 5 Histograms of maximum WSSR values for 300 health condition mismatch cases at cruise

From the histograms generated at various operating conditions, it was found that the largest maximum WSSR value generally increases and becomes an outlier in the distribution as the power setting increases. This tendency can be seen in Fig. 5. After extensive study, it was found that this is due to the fact that health condition mismatches (as large as ±0.5%) can on rare occasions result in quite large mismatches in the state variables and sensor outputs between the engine and the OBEM at high power settings. Under the presence of such large mismatches in sensor outputs, the hybrid Kalman filter performs poorly since its linear component can improve the sensor output matching only to a limited extent.

From the above-noted observation, it would be a reasonable approach to use different threshold values at different power settings in order to achieve effective fault detection performance. The threshold selection in this section is focused on the cruise condition where the power setting is in the intermediate range. At high power settings, such as takeoff and climb, different thresholds must be determined.

As can be seen in Fig. 5, the largest maximum WSSR value becomes a noticeable outlier in the distribution around 70 deg PLA. Similar tendencies were also observed at different flight conditions. Since the nominal power setting is in the range of 60–65 deg PLA during cruise phase, the threshold is set to the value of 1.60. With this threshold value, some margin is available at 65 deg PLA. This threshold value is used for the PLA value up to 69 deg. To further ensure that the threshold violation is due to

[1]Engine health degradation is one of these factors, but its contribution to the cause of false alarms is reduced through the health baseline update. Other examples of nonfault-related factors are customer bleeds, horsepower extractions, and dirt washout from fan and compressors.

Table 4 Minimum detectable sensor bias at 65 deg PLA (numbers in terms of standard deviations)

	Nominal health	Degradation A	Degradation B
XN12	6.4/−6.4	6.9/−4.6	7.4/−3.2
XN25	5.4/−5.4	5.4/−3.9	6.4/−3.0
P25	1.9/−1.9	1.6/−1.8	1.3/−1.6
T25	2.9/−2.9	2.2/−2.8	1.7/−3.3
PS3	2.4/−2.4	1.4/−3.2	0.9/−3.7
T3	3.7/−3.7	3.0/−3.8	2.3/−3.4
T49	4.4/−4.5	4.0/−4.0	3.6/−3.2
T_{amb}	0.4/−0.5	0.3/−0.5	0.2/−0.6
P_{amb}	0.7/−0.6	0.8/−0.5	0.8/−0.3
T2	0.6/−0.6	0.7/−0.4	0.7/−0.3
P2	···	···	···

Table 5 Minimum detectable sensor bias at 69 deg PLA (numbers in terms of standard deviations)

	Nominal health	Degradation C	Degradation D
XN12	6.0/−5.9	6.0/−4.5	1.0/−10.4
XN25	5.9/−6.0	5.5/−4.4	1.1/−7.0
P25	1.9/−1.9	1.6/−1.8	2.6/−0.3
T25	3.0/−2.9	2.7/−2.7	3.3/−0.9
PS3	2.3/−2.3	1.5/−2.7	2.5/−0.4
T3	3.5/−3.5	2.5/−3.8	2.8/−0.8
T49	4.2/−4.2	3.2/−4.3	2.1/−1.6
T_{amb}	0.4/−0.4	0.4/−0.5	0.6/−0.1
P_{amb}	0.7/−0.6	0.7/−0.6	0.1/−1.0
T2	0.6/−0.6	0.7/−0.5	0.1/−0.8
P2	···	···	···

the existence of a fault, the persistency of threshold violation is checked before declaring the fault. It was determined that the threshold must be violated 25 consecutive time steps (0.5 s) to declare fault detection. The threshold value and persistency test are adjusted based on the engineering judgment of the designer, and the performance of the fault detection system will vary with those design factors. In the following, the performance of the fault detection system is evaluated using the threshold value of 1.60.

Performance Evaluation

In the following, the sensor fault detection system is evaluated at a cruise condition. First, the minimum bias that the system can detect for individual sensors is determined at two power settings (PLA=65 and 69) with three different health condition cases. Then, the fault detection system is evaluated to determine if it would generate a false alarm during transient operations by running the engine simulation model through moderate transient scenarios.

Sensor Bias Detection. The value of the detection threshold determined in the previous section was based on the 300 cases of health condition mismatch between the engine and the OBEM. With this threshold value, no false alarm is generated at least for those 300 cases of health condition mismatch. As discussed earlier, decreasing the chance of generating false alarms also decreases the chance of detecting faults. Thus, to evaluate the suitability of the threshold value, the minimum bias that can be detected for individual sensors is determined in this section. If the level of detectable sensor bias is not acceptable, the threshold value must be adjusted accordingly.

The minimum detectable bias is determined for individual sensors by running the engine simulation and the hybrid Kalman filter for 100 s at a steady-state cruise condition. When the fault indicator signal (WSSR) exceeds the detection threshold for 25 consecutive time steps, a fault is detected.

Table 4 shows the minimum bias detected for each sensor at 65 deg PLA at three different health conditions. The bias values are given in terms of measurement standard deviations. For the nominal health case, both the engine and the OBEM are set to the nominal health condition. This case represents an ideal scenario where no health mismatches exist between the engine and the OBEM. For degradation case A, the engine and the OBEM are set to degraded and estimated health conditions, respectively. This case represents a more realistic scenario where health condition mismatches exist. As can be seen in Fig. 5, health condition mismatches cause the WSSR value to vary. The health condition used in case A is the one whose WSSR value represents the 90th percentile value in the 65 deg PLA histogram of Fig. 5; the WSSR values of 270 cases (90%) are less than or equal to the WSSR value at the health condition used in case A. Degradation case B is similar to case A, except that the health condition used in case B results in the largest WSSR value among the 300 cases.

From Table 4, it can be observed that, for the nominal health case, the bias magnitudes that can be detected are symmetric between the positive and negative directions. Such symmetry, however, does not appear in degradation cases A and B where the health condition mismatches exist. The health condition mismatches result in sensor output mismatches between the engine and the OBEM. Consequently, the WSSR value increases and gets closer to the detection threshold as seen in Fig. 5. Therefore, a smaller magnitude bias, when compared to the nominal health case, can be detected if a bias and the corresponding sensor's output mismatch are in the same direction (positive or negative). On the other hand, if a bias is in the opposite direction of the corresponding sensor's output mismatch, the bias counteracts the sensor output mismatch to some extent, and thus it requires a larger magnitude to be detected. From Table 4, it can be observed that the skewness between the positive and negative biases is more prominent in degradation case B than case A.

Table 4 also shows that very small biases are detected for those sensors which indicate the ambient and engine inlet conditions. At first glance, these numbers may lead one to conclude that the fault detection system is too sensitive to biases in those sensors. The reason for such sensitivity is that three sensors (T_{amb}, P_{amb}, T2) define the flight condition for the OBEM and thus a bias in these sensors causes the OBEM to operate at a different flight condition. Since altitude and Mach number can be calculated from those three measurements, it may be necessary to compare the calculated values to measured altitude and Mach number in order to validate those three sensors. The fault detection system was not able to detect a bias in the P2 sensor. This sensor is currently used for the correction of pressure measurements. Since both measured and estimated pressure values are corrected by the P2 value, a bias in this sensor does not increase the residuals.

Table 5 shows the minimum bias detected for each sensor at 69 deg PLA at three different health conditions. This example is similar to the previous case except that the power setting is higher, and thus health condition mismatches can results in the WSSR value closer to the detection threshold. For the nominal health case, both the engine and the OBEM are set to the nominal health condition. For degradation cases C and D, the engine and the OBEM are set to degraded and estimated health conditions, respectively. The health condition mismatch of case C results in the 90th percentile WSSR value in the 69 deg PLA histogram of Fig. 5, whereas that of case D results in the largest WSSR value among the 300 cases.

At the nominal health condition, the fault detection system appears to perform in a consistent manner; the level of detectable bias at this power setting is similar to that of the previous case of the 65 deg PLA, and the detectable bias magnitudes are symmetric between the positive and negative directions. The abovementioned observation, however, does not hold for degradation

cases C and D. Especially for degradation case D where the WSSR value gets very close to the detection threshold, the level of skewness between the positive and negative directions is very prominent. For some of the sensors, smaller magnitude biases, when compared to the nominal health case, can be detected in both positive and negative directions. This gives an impression that it is advantageous to have health condition mismatches so that smaller magnitude biases can be detected. It should be understood, however, that the closer the WSSR gets to the detection threshold due to health condition mismatches, the higher the chance of generating false alarms becomes. From Tables 4 and 5, it is obvious that health condition mismatches and power settings have significant influence on the level of detectable sensor biases. The reader should be reminded that these results show the level of bias that will trigger a fault detection alarm. The detection algorithm will not identify the biased sensor; it only indicates that a fault exists.

Minor Transient Operation for False Alarm Test. The detection threshold used in this paper has been fixed to a specific value for its application during cruise phase. The fixed threshold seems to be applicable as long as the engine operates under steady-state conditions. Since it is desirable to extend the operation of the fault detection system to transient cases, the applicability of the fixed threshold is evaluated in this section by running the engine simulation through minor transient operations. If minor transient operation causes the WSSR value to exceed the fixed threshold, an adaptive threshold, whose value changes as the engine undergoes transients, must be used as was done in Ref. [13] in order to avoid false alarms.

For this evaluation, the engine and the hybrid Kalman filter are run under the following six transient scenarios:

(1) PLA is ramped from 60 deg to 69 deg in 1 s. After steady-state operation, PLA is ramped back to 60 deg in 1 s.
(2) In addition to the PLA ramp in scenario 1, altitude is increased by 305 meters in 10 s. After steady-state operation, altitude is decreased by 305 meters in 10 s.
(3) In addition to the PLA ramp in scenario 1, Mach number is increased by 0.02 in 5 s. After steady-state operation, Mach number is decreased by 0.02 in 5 s.
(4) The same as scenario 1 except that a turbine clearance model is added to the engine model.
(5) The same as scenario 2 except that a turbine clearance model is added to the engine model.
(6) The same as scenario 3 except that a turbine clearance model is added to the engine model.

Each transient scenario is 200 s long, and the time histories of three inputs (PLA, altitude, Mach number) are shown in Fig. 6. Scenarios 4–6 are a repeat of scenarios 1–3 except that a turbine clearance model is added to the engine model. Since it is unlikely that the real engine is perfectly modeled in the OBEM, there will be model mismatch, other than health condition mismatches, between the real engine and the OBEM. In scenarios 4–6, a turbine clearance model, which represents the turbine clearance dynamics with high fidelity, is added only to the engine model representing a real engine in order to introduce unknown dynamics. The presence of the turbine clearance model introduces sensor output mismatches between the engine and the OBEM during transient operations.

The engine simulation and the hybrid Kalman filter were run through the six transient scenarios at five health conditions: nominal health and degradation levels used in cases A–D in Tables 4 and 5. Table 6 shows the result of this simulation test. The symbol "●" indicates that no false alarm was generated, "×" indicates that a false alarm was generated. Figure 7 shows the time histories of the WSSR for transient scenario 2 at the nominal health (solid line) and degradation D (dashed line).

As was noted in Fig. 5, the WSSR gets very close to the detection threshold at degradation case D with 69 deg PLA. The minor transient of scenario 2, in which both PLA and altitude are ramped, causes the WSSR to exceed the threshold temporarily but long enough (at least 25 consecutive time steps) to generate a false alarm. In order to avoid false alarms, the threshold value or the period of persistency check must be adjusted as the engine undergoes transient operations.

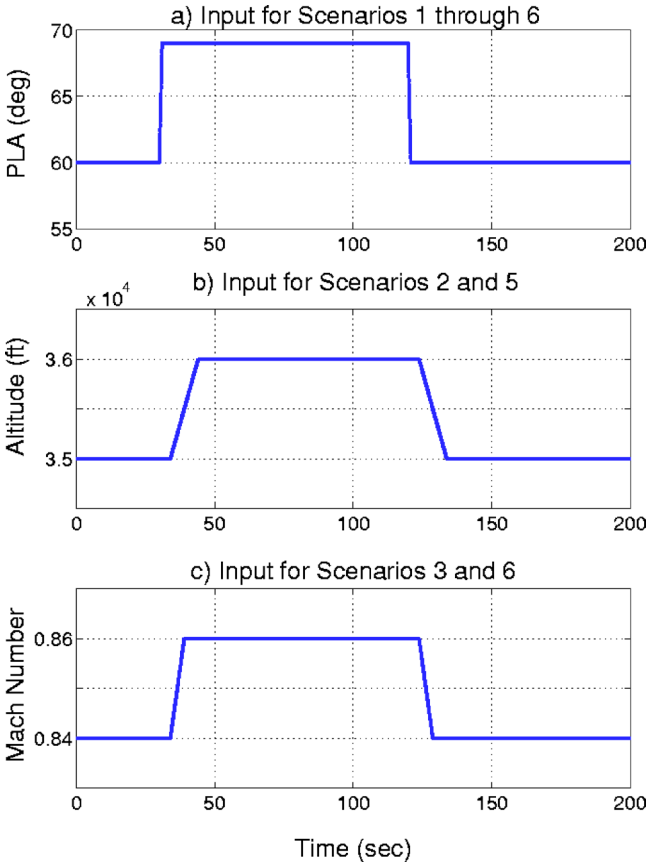


Fig. 6 Time history of transient inputs

Another observation made during the transient test is that the turbine clearance model caused an increase in the sensor output mismatch between the engine and the OBEM at the nominal health and degradation case D. However, in other degradation cases, the turbine clearance model caused a reduction in the sensor output mismatch. Similar to some of the sensor bias cases, it appears that the turbine clearance model counteracted health condition mismatch, resulting in better sensor output matching between the engine and the OBEM. Thus, the presence of unknown dynamics in the engine does not mean that the estimation performance of the Kalman filter always gets worse.

Discussion

The unique structure of the hybrid Kalman filter possesses advantages over conventional Kalman filter approaches and is well suited for application to in-flight diagnostics. In the following, some benefits of the hybrid architecture are discussed.

One obvious benefit of the hybrid architecture is that the reference health baseline of the hybrid Kalman filter can be updated to the health condition of the degraded engine in a relatively simple manner: by feeding the estimated health condition values to the OBEM. This update process is much simpler than for the case of the pure piecewise linear Kalman filter (PLKF) approach. Without the baseline update, the fault detection system loses its diagnostic effectiveness as the real engine degrades over time. To evaluate the significance of the baseline update, the hybrid Kalman filter

Table 6 False alarm test with minor transient scenarios (●: no false alarm, ×: false alarm)

		Transient scenario					
		1	2	3	4	5	6
Health Condition	Nominal	●	●	●	●	●	●
	Degr. A	●	●	●	●	●	●
	Degr. B	●	●	●	●	●	●
	Degr. C	●	●	●	●	●	●
	Degr. D	●	×	●	●	×	●

was applied to the 300 degraded engines, used in the previous sections, without going through the health baseline update. When the reference baseline of the hybrid Kalman filter was fixed to the nominal health condition, the fault detection system misinterpreted degradation as a fault in 283 engines out of 300 (94.3% false alarm rate). Obviously, such a high false alarm rate is not acceptable, and thus the baseline update is a necessary step.

Another benefit is that the hybrid Kalman filter combines the advantages of the constant gain extended Kalman filter (CGEKF) approach [8,14] and the PLKF approach [7]. The advantage of the CGEKF over the PLKF is its capability to capture the nonlinearity of engine operation under the influence of faults. For instance, when an engine experiences a fault, the control system adjusts the actuator positions to meet its objective, such as maintaining fan speed at the commanded value. Due to such control adjustments, the engine moves to a new operating condition which may be a significant deviation from the condition before the fault occurrence. Such nonlinear engine operation due to closed-loop control effects in the presence of fault can be captured by the CGEKF and the hybrid Kalman filter but not by the PLKF. Although the CGEKF approach has such an advantage over the PLKF approach, it also has disadvantages in other areas. As noted in Ref. [14], the numerical stability of the CGEKF may not be as robust as that of the PLKF. Since the nonlinear engine model of the CGEKF receives feedback signals (residuals multiplied by a Kalman gain matrix), large residuals may drive the nonlinear engine model out of the range that the model was designed for. If this happens even for a short period, the numerical stability of the CGEKF may be lost. On the other hand, in the hybrid Kalman filter approach, the numerical stability of the nonlinear engine model is not influenced by the estimation process since the OBEM does not receive any feedback signals. The OBEM runs as a stand-alone engine simulation, generating state variables and sensor outputs at a given health baseline. Based on the information provided by the OBEM, the Kalman filter algorithm is processed using the piecewise linear state-space models. As such, the hybrid Kalman filter possesses the numerical stability of the PLKF approach and also the nonlinear estimation capability of the CGEKF approach.

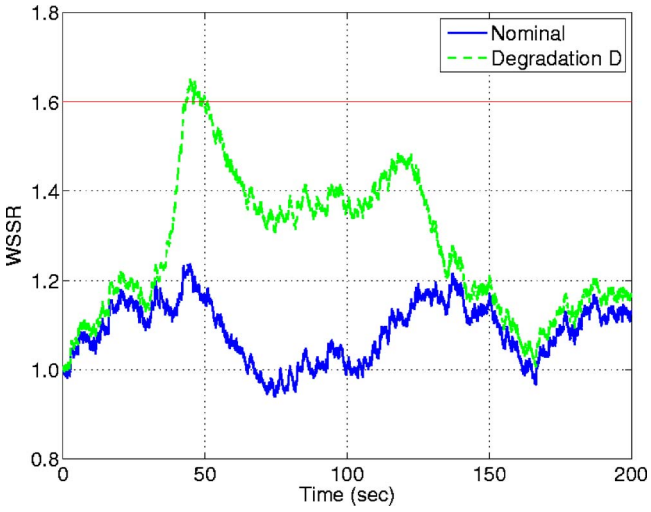


Fig. 7 WSSR responses for transient scenario 2

Finally, the hybrid Kalman filter approach can be easily expanded to a bank of Kalman filters for its application to fault isolation [6,7]. In the hybrid architecture, only the linear component of the filter (A, C, K) must be expanded while keeping only one OBEM. By combining one OBEM and multiple piecewise linear models, each of which is designed based on a unique fault hypothesis, a bank of hybrid Kalman filters can be formed. Therefore, this level of expansion is similar to the case of the pure PLKF approach.

Conclusion

The hybrid Kalman filter approach was developed for its application to in-flight sensor fault detection. The hybrid Kalman filter has a unique architecture which is composed of a nonlinear OBEM and piecewise linear state-space models. In this hybrid architecture, the OBEM functions as an integral part between the off-line and real-time diagnostic systems; it operates at a reference health baseline specified by a trend monitoring system and provides information needed to process the Kalman filter algorithm for in-flight diagnostics. Because of this integration, the in-flight sensor fault detection system does not need to deal with engine health degradation by itself.

To validate this approach, the in-flight sensor fault detection system was evaluated in a simulation environment using a nonlinear model of a large commercial aircraft engine. Its performance was evaluated at multiple steady-state operating points at a cruise condition and also evaluated over minor transients. The performance of the sensor fault detection system was satisfactory in general; however, further study is needed in some areas to improve its capability. It was found that a fixed threshold is not good enough to cover the range of operation that an engine may undergo during flight. For some cases of health condition mismatch between the engine and the OBEM, the value of the fault indicator signal became large at high power settings. In order to avoid false alarms, the threshold must be adjusted according to the power setting and also adjusted during transients. Therefore, the utilization of an adaptive threshold is needed to improve the overall capability of the in-flight sensor fault detection system.

Acknowledgment

This research was funded by the NASA Aviation Safety and Security Program as a task under the Propulsion System Safety Technologies Element.

Nomenclature

BST = booster
CGEKF = constant gain extended Kalman filter
HPC = high-pressure compressor
HPT = high-pressure turbine
LPT = low-pressure turbine

OBEM = on-board engine model
P2 = engine inlet pressure
P25 = HPC inlet pressure
P_{amb} = ambient pressure
PLA = power lever angle
PLKF = piecewise linear Kalman filter
PS3 = combustor inlet static pressure
T2 = engine inlet temperature
T3 = combustor inlet temperature
T49 = LPT inlet temperature
T_{amb} = ambient temperature
TMHS23 = BST metal temperature
TMHS3 = HPC metal temperature
TMHS41 = HPT nozzle metal temperature
TMHS42 = HPT metal temperature
TMHS5 = LPT metal temperature
TMSHBC = combustor case metal temperature
TMHSBL = combustor liner metal temperature
VBV = variable bleed valve
VSV = variable stator vane
WF36 = fuel flow
WSSR = weighted sum of squared residuals
XN12 = fan speed, measured
XN25 = core speed, measured
XNH = core speed, actual
XNL = fan speed, actual
e = environmental parameter vector
h = health parameter vector
h_{ref} = reference health condition vector
u_{cmd} = control command vector
v = sensor noise vector
x = state variable vector
y = sensor output vector (controls/diagnostics)
z = sensor output vector (ambient/engine inlet)

References

[1] Doel, D. L., 1994, "TEMPER-A Gas Path Analysis Tool for Commercial Jet Engines," J. Eng. Gas Turbines Power, **116**, pp. 82–89.
[2] Volponi, A. J., 1994, "Sensor Error Compensation in Engine Performance Diagnostics," ASME Paper No. 94-GT-58.
[3] Kobayashi, T., and Simon, D. L., 2005, "Hybrid Neural-Network Genetic-Algorithm Technique for Aircraft Engine Performance Diagnostics," J. Propul. Power, **21**, pp. 751–758.
[4] Mathioudakis, K., and Romessis, C., 2004, "Probabilistic Neural Networks for Validation of On-Board Jet Engine Data," Proc. Inst. Mech. Eng., Part G: J. Aerosp. Eng., **218**, pp. 59–72.
[5] Romessis, C., and Mathioudakis, K., 2003, "Setting Up of a Probabilistic Neural Network for Sensor Fault Detection Including Operation With Component Faults," J. Eng. Gas Turbines Power, **125**, pp. 634–641.
[6] Kobayashi, T., and Simon, D. L., 2003, "Application of a Bank of Kalman Filters for Aircraft Engine Fault Diagnostics," ASME Paper No. GT2003–38550.
[7] Kobayashi, T., and Simon, D. L., 2005, "Evaluation of an Enhanced Bank of Kalman Filters for In-Flight Aircraft Engine Sensor Fault Diagnostics," J. Eng. Gas Turbines Power, **127**, pp. 497–504.
[8] Kobayashi, T., Simon, D. L., and Litt, J. S., 2005, "Application of a Constant Gain Extended Kalman Filter for In-Flight Estimation of Aircraft Engine Performance Parameters," ASME Paper No. GT2005–68494.
[9] Rausch, R., Viassolo, D. E., Kumar, A., Goebel, K., Eklund, N., Brunell, B., and Bonanni, P., 2004, "Towards In-Flight Detection and Accommodation of Faults in Aircraft Engines," AIAA Pap. No. 2004–6463.
[10] Rausch, R. T., Goebel, K. F., Eklund, N. H., and Brunell, B. J., 2005, "Integrated In-Flight Fault Detection and Accommodation: A Model-Based Study," ASME Paper No. GT2005–68300.
[11] Volponi, A. J., 1999, "Gas Turbine Parameter Corrections," J. Eng. Gas Turbines Power, **121**, pp. 613–621.
[12] Luppold, R. H., Roman, J. R., Gallops, G. W., and Kerr, L. J., 1989, "Estimating In-Flight Engine Performance Variations Using Kalman Filter Concepts," AIAA Pap. No. AIAA-89–2584.
[13] Merrill, W. C., DeLaat, J. C., and Bruton, W. M., 1988, "Advanced Detection, Isolation, and Accommodation of Sensor Failures-Real-Time Evaluation," J. Guid. Control Dyn., **11**, pp. 517–526.
[14] Sugiyama, N., 2000, "System Identification of Jet Engines," J. Eng. Gas Turbines Power, **122**, pp. 19–26.

Dan Martis
Rolls-Royce Canada,
9545 Cote-de Liesse,
Dorval, Quebec, H9P 1A5, Canada
e-mail: dan.martis@rolls-royce.com

Arthur Rowe
Rolls-Royce Civil,
Moor Lane,
Derby, DE24 8BJ, United Kingdom
e-mail: arthur.rowe@rolls-royce.com

Brian Price
e-mail: brian.price@rolls-royce.com

Darryl Edwards
e-mail: darryl.edwards@rolls-royce.com

Vincent Perez
e-mail: vincent.perez@rolls-royce.com

Rolls-Royce Canada,
9545-Cote-de Liesse,
Dorval, Quebec, H9P 1A5, Canada

Marie-Laure Pinquier
6A Rue des Carrières,
Essert, France 90850
e-mail: marielaurepinquier@hotmail.com

Samer Mhana
e-mail: samer.mhana@rolls-royce.com

Nicholas J. Papaconstantinou
e-mail: nick.papaconstantinou@Rolls-Royce.com

Rolls-Royce Canada,
9545 Cote-de Liesse,
Dorval, Quebec, H9P 1A5, Canada

Engine Performance Improvement by Controlling the Low Pressure Compressor Working Line

The new algorithm provides closed-loop control of the LP compressor working line in such a way as to maintain compressor stability and to provide increased power from the engine at the same firing temperature. This method is intended for the Trent 60, an aeroderivative engine designated for power generation and mechanical drive applications. The power benefit is achieved by operating at higher LPC pressure ratio and thereby increasing the inlet air flow and overall pressure ratio of the engine. Since the algorithm controls the working line, the threats to compressor stability related to the working line level are removed (including production scatter, deterioration, and fouling) and the required surge margin can be safely reduced, providing a significant benefit in engine performance. The paper presents comparatively the structure of the current and new concepts, the main features of the controller, and stresses the improved accuracy and reliability of new algorithms. The performance benefit is then assessed; the increase in power is about 3% at ISO, sea level conditions and varies with ambient temperature.
[DOI: 10.1115/1.2718229]

Keywords: compressor, variable geometry optimization, working line, closed-loop control, performance improvement

Introduction

The new concept is intended to improve the performance of the Trent 60, a Rolls-Royce engine designated for ground applications: power generation and mechanical drive.

Trent-60 power unit is currently operated in simple cycle, generating electric power at a high overall efficiency or in combined cycle with a steam turbine powered by the exhaust heat recuperated within a steam generator [1].

Nominal performance parameters at ISO, sea level, uninstalled conditions, simple cycle, Canadian natural gas, are presented in Table 1 as per Ref. [2].

The core engine modules—high and intermediate pressure systems—are common with the aero Trent 800 series (see Fig. 1).

In order to enhance the engine performance and operability for industrial applications and to comply with emission regulations, this product has been developed with the following features [3]:

- dry low emissions (DLE) combustion system
- two-stage booster (replacing the fan)
- modified low pressure turbine driving the booster and the load (hot-end drive)
- for power generation applications, the low pressure spool rotational speed is constant, synchronous with electric generator (grid frequency: 50 or 60 Hz)
- the low pressure compressor is assisted by fully modulated variable bleeds and inlet guide vanes in order to maintain stable operation.

The low pressure compressor rotational speed needs to comply with various power-speed laws, depending on the specific application. The concept described in this paper optimizes the working line position by using the compressor variable inlet guide vanes in

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received May 5, 2006; final manuscript received November 13, 2006. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2005: Land, Sea and Air (GT2005), June 6–9, 2005, Reno, NV. Paper No. GT2005-68729.

Table 1 Overall performance of power generation Trent 60

Trent 60-Version	50 Hz	60 Hz
Shaft power (MW)	51.92	52.16
Shaft efficiency (%)	42.2	42.0
Exhaust mass flow (kG/S)	152.7	155.6
Exhaust temperature (°C)	444	440

order to obtain the maximum shaft power benefit at the same firing temperature.

Current Method for LPC Operability Control

In power generation, the low pressure spool runs at synchronous speed in order to provide the appropriate grid frequency. Therefore, in order to maintain the working line within a stable and safe domain, the low pressure compressor (LPC) needs to be assisted by fully modulated variable geometry—bleed-off valves (BOV) and variable inlet guide vanes (VIGV).

Figure 2 displays a generic characteristic of a LPC running at constant speed. The corrected speed lines define the same physical speed at various LPVIGV angles. The more open the VIGV, the higher the pressure ratio and the flow capacity.

The LPC pressure ratio is scheduled to increase with the power demand and, given the LPC synchronous speed, the LP BOV and VIGV are modulated accordingly.

Between idle and LPBOV closing point, the working line (red solid line) runs along the speed line and is controlled in open loop by the LPBOV doors position. The bleed doors position is scheduled against corrected output power (Sh_Pwr/δ) and corrected speed (NL/RT).

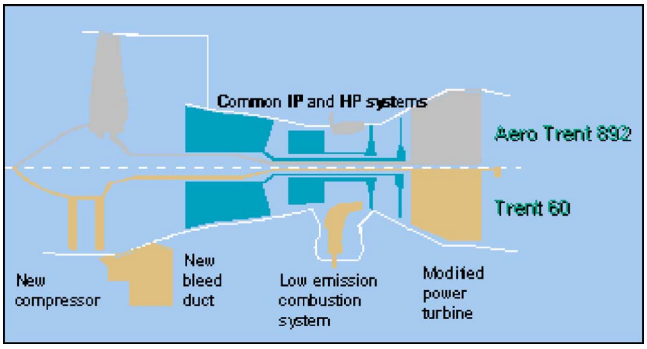


Fig. 1 Comparison between aero and industrial Trent

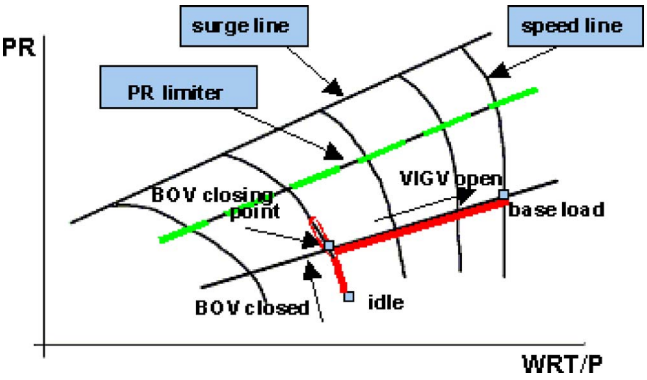


Fig. 2 Generic characteristic of a LPC running at constant speed; the compressor is assisted by fully modulated BOV and VIGV

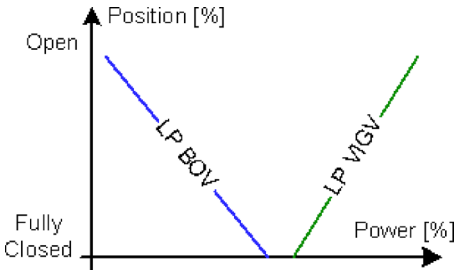


Fig. 3 LPBOV and LPVIGV schedule versus power for a new and clean engine

Above the LPBOV closing point and up to base load power, the LPC working line (W/L) continues to be controlled in open loop via LPVIGV angle against the same parameters: corrected power (Sh_Pwr/δ). This is also shown in Fig. 2.

In order to avoid entering the stall/surge domain, the LPC is provided with a pressure ratio limiter. Once the W/L level tends to exceed the limiter level, the engine control system (ECS) modulates the LPBOV doors opening to maintain the W/L at the limiter level. The algorithm inside the ECS compares the calculated LPC pressure ratio (PR) with the PR limiter level (green dotted line in Fig. 2) and commands the LPBOV actuators to open the doors when the calculated PR tends to exceed the limit.

Summarizing, the current logic comprises two actuation system demands, both scheduled in an open-loop manner against corrected power and speed. Between idle and midpower range, the LPBOV close gradually along the speed line; then the LPVIGV are scheduled to increasingly open from the LPBOV closing point up to base load power (Fig. 3).

Along the entire power range, the controller limits the engine operation on LPC pressure ratio limiter.

However, above the LPBOV closing point, as the working line of the engine degrades (moves up) a risk remains that the LP compressor pressure ratio limiter may be reached. In this case the LP BOV will open to a sufficient extent to maintain engine operation along this limiter. Hence, at high power conditions, it is possible that both the LP BOV and the LP VIGV actuation systems will be modulating (Fig. 4).

LPC Pressure Ratio Assessment

The current algorithm calculates LPC pressure ratio and compares it with the PR limiter map. The calculated LPC pressure ratio (P23SYN/P22) is based on several input parameters and a number of look-up maps derived from the LP and IP compressors characteristics.

Input parameters: P22, P24S, T22, NI.

Look-up maps used:

- LPC map; T24/T22 versus P23/P22 derived from the nominal compressor efficiency map
- IPC map; WRT/P24 versus NI/RT24 derived from IPC nominal working line level map

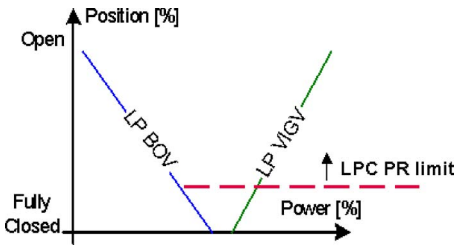


Fig. 4 LPBOV and LPVIGV overlapping schedules-deteriorated engine

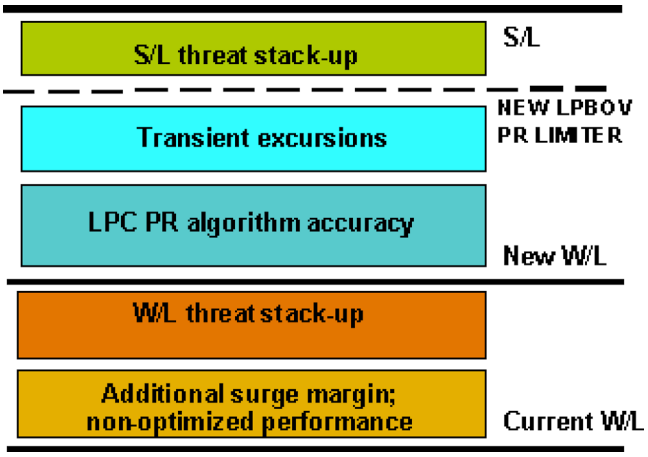


Fig. 5 LP compressor generic surge margin stack-up with and without LPVIGV optimization

- Q curve: P24/P24S versus WRT/P24
- intercompressor duct (ICD) pressure loss map: P23 versus P24

The algorithm calculates iteratively P23/P22 using the following steps:

- guess P23/P22 initial value
- calculate T23/T22 versus P23/P22 (LPC efficiency map)
- calculate NI/RT24 (measured NI and T24=T23)
- calculate WRT/P24 from NI/RT24 (IPC W/L map)
- calculate P24/P24S from WRT/P24 (Q curve)
- calculate P24 having P24S measured
- calculate P23 from P24 (ICD pressure loss map)
- calculate P23/P22, where P22 measured
- calculate error comparing P23/P22 calculated with P23/P22 guessed
- restart the loop until the error within threshold

The above presented algorithm is able then to calculate the current LPC pressure ratio within an accuracy depending on input parameters measurement precision and the approximated look-up maps. It is noted also that the P23SYN accuracy level is associated with the scatter/degradation of the components compared to the nominal maps. The accuracy stack-up has been assessed and introduced as a component in the available surge margin (Fig. 5). These errors are considered as scatter, and their statistical summation induce a S/M reduction no greater than 3%.

Current LPC Surge Margin and Working Line Position

The required surge margin (S/M) for a compressor is made up of a statistical sum of various threats:

- engine to engine variation (production scatter)
- engine deterioration and fouling
- variables mal-scheduling
- transient excursions above the steady state working line
- instrumentation accuracy, etc.

These threats erode the surge margin progressively, by increasing the working line (W/L) level and by lowering the surge line (S/L) level. It is expected that during the life of the engine, the LPC W/L will move up progressively as the engine deteriorates.

With the current algorithm, the W/L level of a new and clean engine is designed to be relatively low compared to the limiter. This is to prevent the W/L from hitting the limiter and therefore to prevent the engine from operating with LPBOV open, when the engine is deteriorated. The margin between W/L and limiter accounts for the accuracy of the W/L location and for the increase of W/L level due to engine deterioration. The level of W/L is therefore set relatively low, which results in nonoptimum LPC pressure ratio and efficiency; as a consequence, the overall cycle performance is also affected.

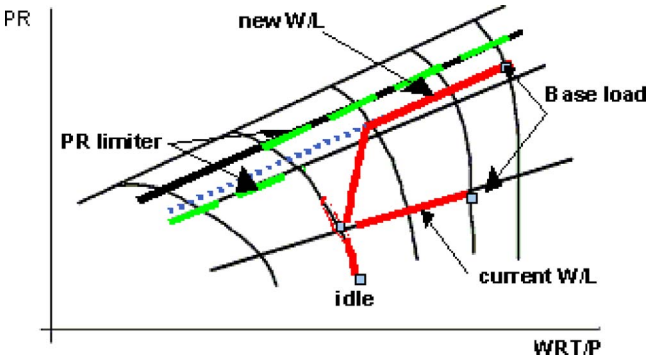


Fig. 6 LPC working line level with optimized LPVIGV position

Description of the New Method

In order to increase the W/L level and to improve the LPC operability above the LPBOV closing point, the new method replaces the two algorithms above with one closed-loop control on pressure ratio, based on LPVIGV position and NL/RT [5].

Above the LPBOV closing point, the LPC W/L is controlled and maintained at the highest safe level during the entire engine life by modulating only the LPC VIGV. Since the working line is maintained at a constant level, the threats related to the working line are removed (production scatter, deterioration, fouling), and the required surge margin (S/M) can be safely reduced. The only threats remaining on the stack-up are related to S/L level deterioration, transient excursions, and LPC PR accuracy degree (Fig. 5).

The main advantages of the new method are:

1. Due to the constant level of working line throughout engine life, the allowance of W/L pressure ratio increase due to production scatter, deterioration, and fouling can be eliminated and the controlled surge margin can be reduced. Therefore, the LPC pressure ratio may be increased, providing benefits in engine performance. This increased LPC pressure ratio improves the cycle by forcing more air into the core engine and fundamentally increasing the power output capability of the machine at a firing temperature (T41).
2. The LPC working line level is independent of output power. Therefore the new algorithm eliminates the risk of VIGV angle closing due to power deterioration and consequently improves performance retention.
3. The new algorithm eliminates the risk of the LPC pressure ratio limiter becoming active as the engine deteriorates. Therefore the efficiency loss due to LPBOV opening is eliminated.

Figure 6 presents comparatively the increase in W/L level between the LPBOV closing point and base load power. The new logic is enabled right after the point where LPBOV close and then the engaging sequence brings the running point up to the new W/L level. As mentioned before, the new logic operates at the highest safe PR level in order to maximize the engine mass flow, overall pressure ratio, and output power. The PR limiter is moved up in order to prevent interference with the PR target and to provide supplementary protection (Figs. 5 and 6). Additionally, to avoid both actuation systems being enabled at the same time, the LP VIGV actuation system is not permitted to control until the LP BOV system has fully closed (Fig. 7).

The new algorithm maintains the LPC W/L level constant throughout the engine life by closed-loop LPC PR control. The

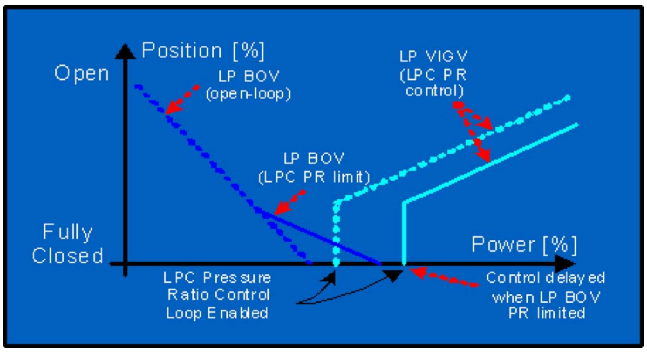


Fig. 7 Improved BOV and VIGV schedule by applying the new logic

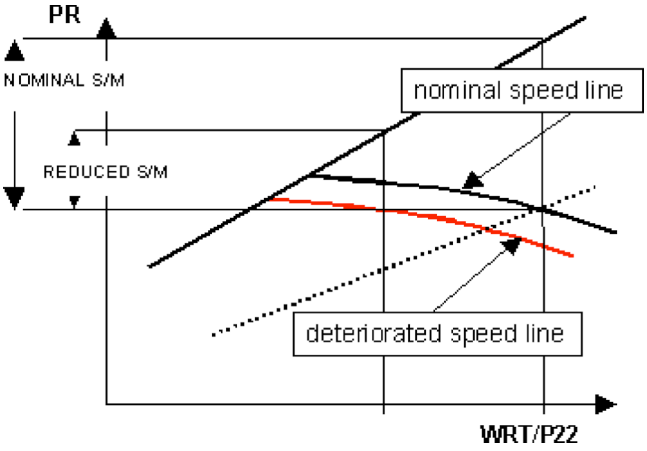


Fig. 9 Speed line drift for a deteriorated engine

controller is displayed in Fig. 8. The ECS look-up map calculates the LPC PR demand versus the current corrected spool speed (NL/RT) and LP VIGV angle. The LPC PR feedback coming from the engine (via P23SYN algorithm) is compared with its demand and the error between these parameters (LPC PR error) is used as an input to the LPC PR controller. The LPC PR controller demands a VIGV position and LP VIGV position feedback is attained via closed-loop control. The LP VIGV position feedback is used also as an input for the LPC PR look-up map, where:

Blocks:

A: Lookup map LPC pressure ratio versus NL/RT and LPVIGV angle

B: LPC pressure ratio controller

C: LP VIGV controller

D: LP VIGV actuator

E: Engine

Signals:

S1: NL/RT

S2: LPC pressure ratio

S3: LPVIGV demand

S4: LPVIGV feedback

S5: LPC pressure ratio feedback (P23SYN/P22)

By using such an algorithm, the ECS can maintain the LPC W/L level constant, independent of engine output power. The new algorithm therefore eliminates the "double hit" caused by VIGV angle closing when power drops.

The signal S5 (LPC pressure ratio feedback) is the result of the algorithm presented in the section entitled "LPC Pressure Ratio Assessment."

Further Improvements: Use of LPC Inlet Flow

The LPC working line control has NL/RT as an input. The speed lines on the LPC characteristic are relatively "flat" (similar to a fan—see Fig. 9); therefore, even though NL and T22 are directly measured parameters, a LPC PR map based on speed lines can create a relatively big uncertainty with effect on surge margin reduction.

In addition, the compressor deterioration and fouling reduce the flow capacity at a corrected speed. Given the flat speed lines, the shift toward the surge line is significant (Fig. 9). This must be accounted as a surge margin "threat" and therefore serves to reduce severely the target W/L level and the output power benefit.

One way of addressing this problem is to use LPC inlet flow function (WRT/P22) instead of corrected speed (NL/RT) as an input signal S1. This way, the error associated to the flow capacity slippage against the corrected speed can be avoided and the P23SYN accuracy stack-up can be reduced.

Therefore, this parameter has been introduced as part of the new logic, replacing NL/RT.

However, WRT/P22 is a derived parameter based on nominal LPC and IPC maps. Although the P23SYN accuracy increases, the drift of LPC and IPC maps will still have a certain effect on the final results.

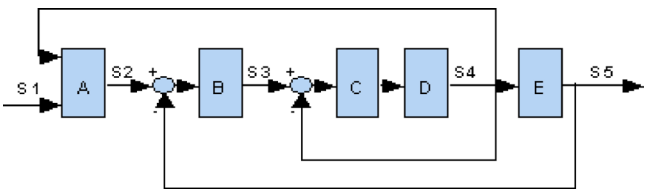


Fig. 8 New controller with closed-loop on LPC pressure ratio

Further Improvements: Direct Use of LPC Outlet Static Pressure

Although using WRTP22 is a better choice, this method depends also on the P23SYN algorithm. Therefore, the level of accuracy is still relatively low and has a slow response. This risk needs to be addressed by lowering the target W/L level in order to account for the algorithm scatter/drift.

A more direct way of controlling LPC W/L is by using the ratio of LPC exit static pressure (in fact IPC inlet static pressure) to inlet total pressure (P24s/P22). In comparison with the current method, which uses LPC total pressure ratio (P23SYN/P22), the new control loop is simpler, faster, more accurate and reliable. A compressor characteristic based on outlet static to inlet total pressure ratio is essentially similar to a characteristic based on total pressure ratio.

There are a number of benefits to controlling against the measured static pressure directly:

- the static pressure changes more steeply than the total pressure as surge is approached, making the control less sensitive to measurement of pressures; this happens because the outlet flow function, and hence Pt/Ps, falls as the compressor is throttled;
- the direct control (P24s/P22) will remove a lot of complexity and uncertainty from the control; the uncertainty stack-up will include only the measurements and the map degradation due to deterioration and fouling.

Figure 10 displays a generic comparison between the two characteristics: a typical total pressure ratio and a static-to-total pressure ratio. It is obvious that, by making the speed lines steeper, the surge margin increases significantly, with direct effect on the LPC stability control. Static-to-total surge margin (DPRs) is more than twice the total-to-total surge margin (DPRt)—Fig. 10.

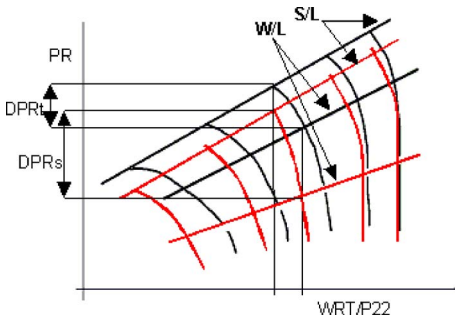


Fig. 10 Static-to-total pressure ratio characteristic compared to total pressure ratio characteristic

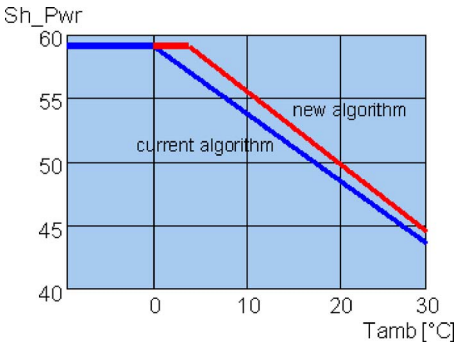


Fig. 11 Power benefit for Trent 60–50 Hz

So, using the static to total characteristics roughly doubles the pressure change between the operating point and surge, which will make accurate control easier and less risky.

The use of LPC static-to-total pressure ratio ensures a better accuracy and a faster response, which might be cashed in as an extra increase in working line level.

Operating the Trent 60 LPC With Reduced Surge Margin

One critical aspect to be addressed was to ensure that the Trent 60 operation with LPC reduced surge margin was satisfactory.

The Trent 60 fleet has already run a significant number of hours with the current logic. Parts of these engines are already running close to PR limiter; sometimes, for deteriorated and heavily fouled engines, the limiter level has been reached for short period of time—either as steady state or transient maneuvers. Therefore, the current experience in running the engine with reduced LPC surge margin provides the certainty that there is no perceptible risk of stall/surge due either to instrumentation failure/drift or to hardware distortion.

In order to increase the engine protection against LPC stall/surge, the existing LPBOV-assisted PR limiter is still used along with the new logic. Its position has been moved up, above the PR target, right below the surge line deteriorated level; once this limiter is hit, the LPBOV will open and the engine will disengage the new logic. Therefore, LPBOV can be regarded as a protection system in the improbable case of LPVIGV failure.

When the new logic will be ready to operate with the LPC static-to-total pressure ratio algorithm, then the P23SYN algorithm will be maintained and run in background. The cross check between the two methods will make the LPC pressure ratio target effectively high integrity; and will allow a comprehensive health monitoring of the engine front end.

Performance Benefit: Comparison

This section presents the benefit in shaft power versus ambient temperature at sea level, uninstalled conditions, Canadian natural gas. For the Trent 60, the power is flat rated at 60 MW for cold days, up to the kink point (around 0°C). Above the kink point, where the T41 limit is reached, the power reduces with the ambient temperature.

By introducing the new logic, the output power at a constant T41 increases by 2.5–3% within an ambient temperature range between 0 and 15°C. The biggest benefit occurs around the kink point (~3%). At 30°C T_{amb}, the extra power is about 2.5% for 50 Hz version (Figs. 11 and 12) and 1.5% for 60 Hz version (Figs. 13 and 14). The difference in benefit between the two engine versions is due to the difference in LPC hardware and the default LPVIGV position related to the current logic.

At constant power, the engine with optimized LPVIGV will run cooler. Above 70% base load power, T41 reduces by 5–10 K, depending on power level and ambient temperature.

The difference in hot end temperatures leads then to an extended engine life—between 10% and 20%, depending on the site-specific conditions.

In conclusion, the benefit can be applied for increasing the guaranteed power level or conversely, for extending the engine life at the same power guarantee.

At constant power, the thermodynamic efficiency of the cycle decreases by about 0.4%, due to the decrease in T41. At constant T41 however, the efficiency is largely unchanged.

Conclusions

By increasing the LPC working line level, the higher pressure ratio forces more air through the core increasing power. Moving from an open to a closed loop control allows an optimized operability of the LPC.

By using this novel control algorithm to raise the LPC W/L level, the Trent 60 can provide extra output power with no associated design change or loss of life. The power increase is about 3% between 0°C and 15°C T_{amb}, with the biggest power benefit occurring around the kink point.

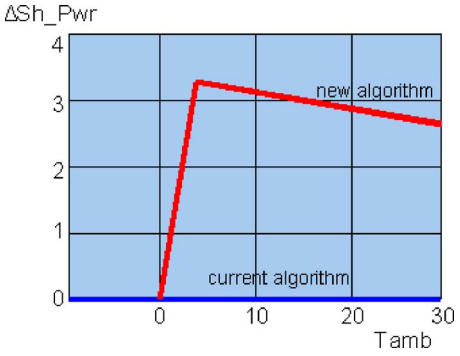


Fig. 12 Power benefit in percents for Trent 60–50 Hz

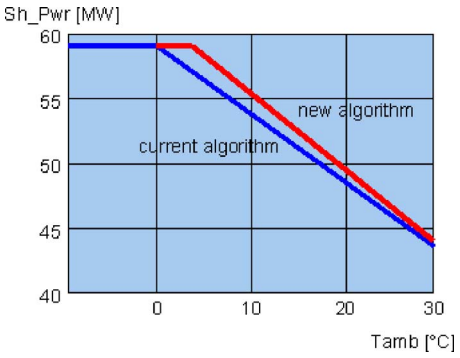


Fig. 13 Power benefit for Trent 60–60 Hz

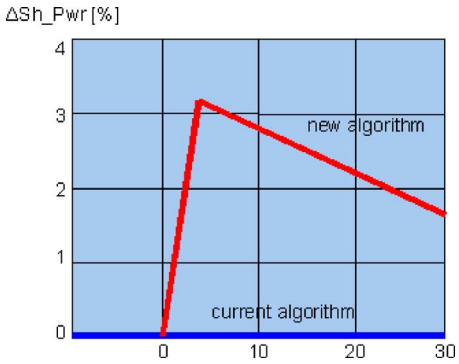


Fig. 14 Power benefit in percents for Trent 60–60 Hz

At constant power, the engine runs cooler and T41 decreases between 5 K and 10 K, depending on the power level and ambient temperature. The engine hot end life extends accordingly, between 10% and 20%.

In addition, as a further improvement, the working line can be controlled directly using outlet static pressure with a significant increase in LPC surge margin; this way, the algorithm accuracy and responsiveness will improve as well.

Acknowledgment

Our gratitude to Chris Barkey, John Uings, Tomas Scarinci, Steve Richards, Kostas Stavrianos, Carl Carson, Robert Ouellette, Lynn Stewart, and Richard Tunstall at Rolls-Royce Energy Business, for their continuous support. Our deepest thanks go to Phil Walsh and Steve Brown, at Rolls-Royce UK, for their technical support and useful advice. Our gratitude to Professor Pericles Pilidis, at Cranfield University, for his continuous technical and moral support.

Nomenclature

BL = base load power
DPRs = delta pressure ratio between surge line and working line, respectively, both defined as outlet static to inlet total pressure ratio
DPRt = delta pressure ratio between surge line and working line, respectively, both defined as outlet total pressure to inlet total pressure
ECS = engine control system
HPT = high pressure turbine
ICD = intercompressor duct, between LPC and IPC
IPC = intermediate pressure compressor
LPC = low pressure compressor
LPBOV = LPC bleed off valves located in the ICD
LPVIGV = LPC variable inlet guide vanes
NI = IPC rotational speed
NI/RT = NI/$\sqrt{}$T—IPC corrected speed
NL = LPC rotational speed
NL/RT = NL/$\sqrt{}$T—LPC corrected speed
P = pressure (psi)
P23SYN = synthesized (calculated) LPC outlet pressure
Ps = static pressure (psi)
PR = compressor pressure ratio
Sh_Pwr = shaft power (MW)
S/L = surge line
T_{amb} = ambient temperature (°C)
T = temperature (°K)
W = mass flow rate (kg/s)
W/L = working line
WRT/P = $W^*\sqrt{}T/P$—compressor corrected inlet flow

Engine Stations [4]:

22 = LPC inlet
23 = LPC exit
24 = IPC inlet/ICD outlet
41 = HPT rotor inlet temperature
δ = Pamb/Psls—ratio between local ambient pressure and sea level standard pressure

References

[1] Barkey, C., Richards, S., Scarinci, T., and Kotsiopriftis, P., 1999, "Industrial Trent: Combustion and Other Technologies," XIV International Symposium on Air Breathing Engines, Florence, Italy.
[2] *Gas Turbine World Handbook*, 2003, Pequot Publishing, CT.
[3] Myers, G. A., and Jackson, A. J. B., "Development of the Trent Econopac," June 13–16, 1994, ASME 94-GT-446, The Hague, Netherlands.
[4] SAE, 1974, "Gas Turbine Engine Performance Station Identification and Nomenclature, Aerospace Recommended Practice," ARP No. 755A, Society of Automotive Engineers, Warrandale, PA.
[5] Martis, D., and Rowe, A., 2004, "System and Method for Controlling the Working Line Position in a Gas Turbine Engine Compressor," Patent Application US2006101826 filed Nov. 12, 2004, published May 18, 2006.

Enhanced Gas Turbine Performance Simulation Using CFD Modules in a 2D Representation of the Low-Pressure System for a High-Bypass Turbofan

Friederike C. Mund
Georgios Doulgeris
Pericles Pilidis

School of Engineering,
Cranfield University,
Cranfield MK 43 0AL, UK

The improvement of performance simulation for gas turbines has been approached in very different ways. In particular for high-bypass turbofans, efforts have been made to investigate radial flow distributions. The aim of the presented study was to combine a conventional characteristics-based performance code using a 2D representation of the fan with 2D representations of the adjoining intake and bypass system. Computational fluid dynamics (CFD) was the chosen tool to generate modules for the intake, bypass duct, and bypass nozzle. This approach required geometry data. A design procedure to generate these components in an axisymmetric meridional fashion and the numerical requirements for the CFD modules were developed. Typical component performances were predicted and the combined use of CFD and the performance code showed that in terms of performance, the inclusion of intake and bypass losses and the radial inlet distribution was worth considering. In particular, however, the required numerical effort was significant. [DOI: 10.1115/1.2364197]

Introduction

The three-dimensional numerical simulation of a gas turbine engine as a whole is leading edge research. The scale of such a computation limits its application to case studies outside the standard design procedure of aeroengines. At the other end of the scale, gas turbine performance simulation based on 0D component characteristics is well integrated into the design process.

For the past two decades the improvement of gas turbine performance modeling by using higher fidelity component models has been pursued. Leading the field is the NASA numerical propulsion system simulation (NPSS), a multidisciplinary environment using state-of-the-art computing and communication technology for the analysis and design of aircraft engines. NPSS supports "numerical zooming" from 0D, 1D, 2D, and 3D component engine codes in terms of aerodynamics, structural engineering, and heat transfer analyses. In close collaboration with engine manufacturers, 3D Navier–Stokes simulations of the low-pressure system (LPS) including external flow, nacelle, inlet, fan, bifurcated bypass and core inlet, bypass vanes, core inlet guide vanes, mixer, and exhaust nozzle have been performed. These were coupled with the core components, modeled using a 0D performance simulation. To keep such complex models within the scope of industrial application, a solution turn-around time for a single operating point was limited to "overnight solutions" and was facilitated by using parallel computing techniques using in excess of 32 processors [1].

This certainly represents the high end of the scale of current performance modeling and was applied to specific engines in close collaboration with engine manufacturers to facilitate test and geometry data exchange. A stepwise path to achieving a higher order of fidelity for modeling an engine is to replace individual engine components of a 0D performance simulation tool with 2D modules. This reduces computational complexity significantly and depending on the number and type of 2D components, generally can be performed on a single processor or personal computer.

The use of 2D modules recognizes the fact that radial inlet profiles influence the performance of the components. For 0D performance models, no extrapolation mechanisms are available to derive radial profiles at component interfaces. This limitation can only be overcome by starting at the most upstream component, the intake, using a radial modeling approach and following the flow downstream.

A 0D performance model for a two-shaft, high-bypass turbofan with a radial representation of the fan was made available to the authors. As the backbone of the performance simulation, it provided radial boundary conditions downstream of the fan and with some modifications was developed to be capable of handling radial inlet profiles. The next logical step was to develop radial representations of the intake, bypass, and bypass nozzle. The chosen numerical tool was CFD, which can be implemented in 2D when assuming axisymmetry. Geometrical data, in this case the meridional cross sections of the components, were required but, in an academic environment, detailed engine data are generally unavailable. The simplification to axisymmetric geometries reduced the extend of the data required and enabled the data set to be generated using basic design principles.

Fan Modeling

A component of particular importance for radial modeling is the fan, especially for high-bypass turbofans when the hub-to-tip ratio of the fan is relatively low and the flow parameters vary considerably across the span. As the fan delivers compressed air to two downstream systems, the core and the bypass, the inlet pressures

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received October 1, 2005; final manuscript received February 1, 2006. IGTI Review Chair: R. S. Abhari. Paper presented at the ASME Turbo Expo 2006: Land, Sea, and Air (GT 2006), Barcelona, Spin, May 8–11, 2006, Paper No. GT2006-90440.

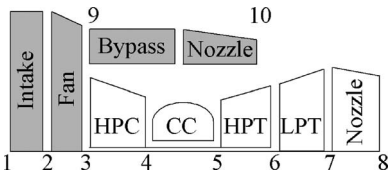


Fig. 1 Performance model for a two-shaft high-bypass turbofan

and temperatures to these systems may differ considerably. The fan therefore provides a strong incentive for an extension to radial modeling.

Riegler et al. [2] presented a fan model for a three-stage low-bypass turbofan where the radial flow profiles at the exit of the fan were derived from rig data. It was shown that the bypass ratio influenced the fan map parameters and a correcting algorithm was presented accounting for this effect. Based on the assumption that the splitter was located far downstream of the fan exit, the suggested approximation would deteriorate with closer proximity of the splitter to the fan exit.

Curnock et al. [3] used a similar fan model for high-bypass turbofan applications. The radial distribution of the pressure ratio, efficiency, and mass flow of the fan were stored in "profiled" characteristics. Notional radial profiles were generated, not specific to a particular fan. However, the artificial profiles were considered to be realistic for a high-bypass turbofan. Previous investigations by Marshall [4] showed that fan profiles were independent of the downstream flow split. Deviations in engine performance of up to 2% in thrust, 1% in specific fuel consumption (SFC), 3% in bypass ratio (BPR) and 8% in core pressure ratio were found when using the 2D representation in comparison with values appropriate to the 0D fan characteristic. A further investigation of the influence of radial inlet distortion on the engine performance indicated an equivalent reduction of thrust and bypass ratio for an intake pressure loss of 1%, along with an equivalent increase of SFC [5].

It was in the nature of both of these models that the fan maps utilized referred to a specific inlet condition. Riegler et al. [2] derived the maps from experimental data, which inevitably referred to the test cell fan-inlet profile. Inlet distortion, however, affects the fan-exit profiles and a coupling of such fan models to an upstream component would require changes of the profiled data and the maps.

The streamline curvature method can handle radial inlet data, but so far it was generally used further into the engine design process, for turbomachinery design and analysis. Recently, its use for a multistage fan coupled with a 0D performance model has been presented for a low-bypass turbofan by Li [6], with encouraging results. Higher benefits can be expected for single-stage fans. An adaptation of the model for a high-bypass turbofan application, however, has not been available to the authors and therefore some adaptations were made to the radial fan model by Curnock et al. [3].

Performance Model of the High-Bypass Turbofan

The performance model by Curnock et al. [3] with a 2D representation of the fan was used for the study. A schematic of the performance model of the investigated two-shaft, high-bypass turbofan with separate exhausts is illustrated in Fig. 1. The shaded components were modeled two-dimensionally, whereas the remaining components employed 0D characteristics. A summary of the performance data at design point at the end of climb is given in Table 1.

The characteristics of the fan were based on undistorted inlet flow. Profiled characteristics were available for the fan flow function, pressure ratio, and efficiency. To reflect radial changes of inlet pressure in the fan-exit profile, the pressure ratio profile was

Table 1 Performance data for the two-shaft high-bypass turbofan at design point

Thrust (N)	20000 N
BPR	3.75
Mass flow	79.04 kg/s
Overall pressure ratio	27.8

adapted as a function of the radial inlet distribution of total pressure. To preserve the 0D maps, the resulting pressure ratio was kept the same. Provided that the incidence angles were unaffected, no changes in radial inlet velocities occurred, allowing for the assumption of the same fan efficiencies. On this basis, the radial distribution of the fan efficiency remained unchanged. A further simplification was the assumption of unchanged radial distributions of the flow function. The overall reduction of engine inlet mass flow due to radial distortion was accounted for by introducing the inlet pressure recovery factor (PRF), defined as the ratio of total pressure at the fan-inlet to the free-stream stagnation pressure of the captured flow, into the simulation.

For each whole-engine simulation, a new set of profiled characteristics was generated for the relevant inlet conditions. This was facilitated by means of a separate Fortran program combined with a MATLAB procedure.

Design of a Generic Propulsion System

CFD calculations require boundary conditions specified in terms of flow parameters and geometrical boundaries. A CFD study therefore inevitably attributes dimensions to the traditional, nondimensional performance prediction. Performance simulation considers component losses and mass flow balances, but geometrical areas can only be derived where additional assumptions were made for the static pressures, as for the choked nozzle-exit condition, for example. Moreover, data are only available at component interfaces, neglecting the streamwise dimension. Such data sets are insufficient for performing two-dimensional CFD simulations and additional data have to be included.

When engine manufacturers base their new engine designs on existing ones, first geometry data are usually in place for preliminary investigations. In an academic environment, detailed geometries are generally not available. The absence of this essential information required the design of a generic geometry data set. The general design procedure for the intake, nacelle, bypass, and bypass nozzle is outlined in the following sections.

Subsonic Intake and Nacelle. The design process for the intake and nacelle geometry compromised the following steps:

- sizing of the intake
- establishment of key internal intake design parameters
- design of the internal lip contour
- design of the diffusion section
- establishment of key external nacelle design parameters
- design of the forebody contour
- layout of the afterbody shape

The air intake limits the mass flow that can enter the engine. Therefore, the intake throat, as the smallest area, has to be large enough to meet the maximum flow demand of the engine. The design procedure for the intake selects the throat Mach number and the downstream diffusion rate to provide the right fan-inlet velocity. A schematic of the dimensions of a subsonic intake and the surrounding engine nacelle is shown in Fig. 2. The key geometric parameters used for intake design are the contraction ratio (CR), defined as the ratio of highlight area to throat area and the gross diffusion ratio (DR_{gr}), defined as the ratio of the circular area defined by the fan tip radius to throat area. The net diffusion ratio (DR_{net}) is based on the net inlet annulus area of the fan,

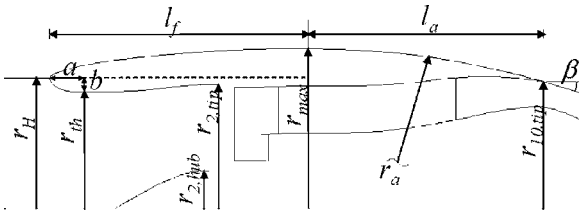


Fig. 2 Key design parameters for a propulsion system

accounting for the presence of the spinner.

It is crucial to size the intake according to the highest engine-flow demand to avoid choking of the intake under any condition within the flight envelope. The nondimensional mass flow rises with altitude and turbine entry temperature (TET) and the sizing point for the intake is usually set at the end of climb. The cruise altitude and aircraft cruise speed are just reached, but the TET is well above the cruise value. Cumpsty [7] suggested a TET of 125 K below the take off condition for the end of climb. The sizing point was therefore modeled at 10670 m altitude, a Mach number of 0.83, and a TET of 1575 K, and the predicted engine mass flow and relevant atmospheric conditions were used to size the fan.

Based on the maximum throat Mach number, DR_{gr}, and the fan hub-to-tip ratio r_{hub}/r_{tip}, the key intake parameters were derived using the "Q function" (see Fig. 3). To allow for future flow requirements, a "flow stretch" of 5% was introduced.

Fans generally have hub-to-tip ratios ranging from 0.3 to 0.6 [8], where high-bypass turbofans tend to have lower values to achieve the highest practical mass flow per unit frontal area [9]. The recommended design ranges for other key intake and nacelle parameters are summarized in Table 2. A lower DR_{gr} was selected, as comparable values were found for commercial engines of the same size. The net internal diffusion ratio was relatively small and a linear streamwise diffusion being chosen. With a generic spinner geometry from a similar engine, the internal inlet contour was designed.

A typical CR of 1.25 was chosen. The entry lip, as one of the most important components of a subsonic inlet, connected the highlight with the throat plane: A super-ellipse curve for improved take-off performance was chosen with a ratio of a/b of 2.0 and an elliptic exponent of 2.0 [10].

The selection of the nacelle forebody geometry is often a compromise to satisfy conflicting requirements. The avoidance of wave drag at high Mach numbers requires a long, thin cowl, but good low-speed spillage-drag characteristics are associated with short, fat cowls [11]. An intake cowl creates local velocities on the external surface that are in excess of the free-stream velocity. With increasing aircraft speed, a region of supersonic flow will develop. This may also be accompanied by flow separation, leading to rapid drag rise. Keeping the surface flow velocities subsonic generally requires diameter ratios of the highlight to the maximum nacelle diameter D_H/D_{max} of around 0.5–0.7. For high-bypass turbofans this ratio usually is in the region of 0.8–0.9. But with careful control of the curvature the formation of a shockwave can

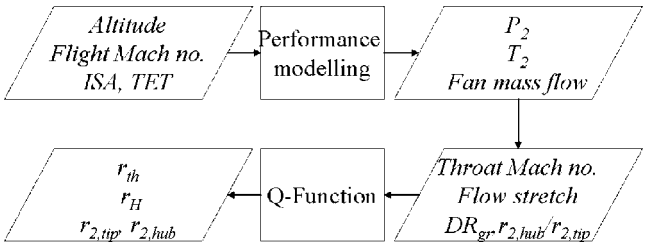


Fig. 3 Procedure to derive key intake design parameters

Table 2 Design guidelines for civil high-bypass turbofan intakes.

			Design range
Fan hub-to-tip ratio	$r_{2,hub}/r_{2,tip}$	0.3	0.3…0.6[a]
Contraction ratio	CR	1.25	1.2…1.35[b]
Diffusion ratio	DR_{gr}	1.125	1.25…1.35[b]
Throat Mach No.		0.72	0.7…0.75[b]
Diameter ratio	D_H/D_{max}	0.815	0.8…0.9[c]
Boat-tail angle	β	16 deg	<16 deg[b]

[a]Sec Ref. [8].
[b]Sec Ref. [14].
[c]Sec Ref. [12].

be weakened and directed towards the thickest section [12].

The NACA 1-series cowl-forebody profiles are based on a profile that can be varied in length and diameter ratio. They have gained wide acceptance for practical use, it being crucial to use an appropriate length ratio for the profile. The intake critical mass flow ratio and drag-rise Mach number can be correlated against the forebody dimensions, based on the empirical relationships developed by Stanhope [13]. A drag-rise margin was introduced by increasing the Mach number at cruise by 0.08 for the chosen layout of the forebody. The resulting forebody profile following the NACA nomenclature can be designated as NACA-1-81-80.

The coordinates of the profile were based on the geometry representation of the NACA-1 series by ESDU [14] for computational fluid dynamics. This analytical approximation of the original data was designed with improved smoothness of the second derivatives to avoid irregularities in the pressure distributions in flow computations.

The afterbody length was designed using the maximum nacelle diameter, which was based on the selected diameter ratio, together with the boat-tail angle and the exit nozzle areas. The aerodynamic design of an axisymmetric afterbody involved fairing the boat tail between the maximum nacelle diameter with the final nozzle. A circular arc profile was used. Less curvature avoided premature drag rise in the expansion and therefore a flow-suction region just downstream of the maximum diameter. The boat-tail angle β was chosen such that boundary layer growth was limited and flow separation avoided [15].

Bypass-Duct and Nozzle. The bypass-duct and nozzle designs were based on the area distribution between the fan exit and bypass-nozzle exit. The bypass-nozzle radii were established with the design of the afterbody and the defined bypass-nozzle exit area for the performance model of the engine. As found in McKenzie [9], the fan-exit tip radius was assumed to be 0.98 of the inlet radius. With recommended bypass-inlet Mach numbers between 0.3 and 0.4 by Walsh and Fletcher [16], a fan-exit hub radius of 1.25 $r_{2,hub}$ resulted for cruise conditions. Downstream of the fan exit, up to the position of the leading edge of the splitter, a constant duct cross-sectional area was assumed.

The area distribution between the fan exit and bypass nozzle is illustrated in Fig. 4. As illustrated with the solid line labeled "Design," a constant area over 70% of the entire length of duct and nozzle was assumed, followed by a converging section with a final area reduction to 64% of the bypass-inlet area. The shape of the converging area reduction was designed using a sixth order polynomial function to achieve a smooth area-change transition between the constant and the converging section. At the nozzle exit, the shape was mainly governed by the external shape of the nacelle—the outer contour of the bypass having to stay within the nacelle outline and a certain distance having to allow for manufacturing clearances. A decreasing slope of the area reduction was necessary to achieve such a configuration.

The inner and outer contour of the bypass duct and nozzle were designed using polynomial functions. The outlines are shown in

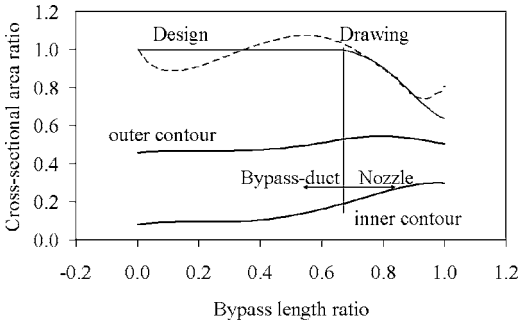


Fig. 4 Streamwise distribution of the cross-sectional bypass area and bypass contours

the lower section of Fig. 4. The area distribution of a scaled drawing of a similar commercial engine was also included (dashed line labeled "Drawing") to demonstrate how misleading such illustrations can be if used for technical investigations.

Setup of 2D CFD Modules

Once a geometry is established, several issues have to be considered:

- shape and size of the computational domain
- grid structure and density
- boundary conditions and
- solver settings

For the present case, a 2D axisymmetric setup with structured meshes was selected. The structured approach was considered to be appropriate as it allowed the mesh to be aligned with the flow in most regions of the domain and thus improved numerical accuracy. The CFD domain included the engine and external flow and was split into two main blocks: The upstream block included the intake and external flow of the forebody and the downstream block included the bypass duct and nozzle and the external flow of the afterbody. These blocks could be detached and calculated individually, but for the present study, the full domain was considered in all cases.

The commercial flow solver "FLUENT" and its preprocessor "GAMBIT" were used to provide the CFD modules [17]. GAMBIT allowed for generating structured grids and featured boundary layer functions for easy application at walls. It was optional to map the boundary layer region and the main grid independently to the computational domain boundaries, which facilitated the generation of a flexible multiblock structure within an individual mesh block.

Figure 5 shows two versions of a 2D mesh block of a cross section of an aeroengine intake duct. In (*a*), a thicker boundary layer was created compared to (*b*). To provide adequate resolution of the boundary layer, the boundary layer meshes were attached to the solid boundaries of the computational block. This body-fitted approach increased numerical accuracy as the flow was well aligned with the mesh. It resulted in a *C* structure around the intake lip. The main mesh in Fig. 5 used an *H* structure where the gridlines led from the left to the right block boundary and the orthogonal gridlines connected the top with the bottom block boundaries. For air intakes operating at forward speed this provided an appropriate mesh structure as the resulting streamlines are closely aligned with the gridlines.

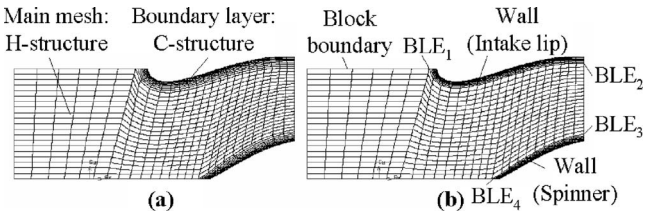


Fig. 5 Structured mesh block using the boundary layer function to create multiblocks

In (*b*), a thinner boundary layer region was chosen and the edges of the boundary layer mesh (BLE1 to BLE4) moved within the overall mesh block. Without GAMBIT's boundary layer function, a decomposition of the domain into subregions would have been required to apply structured meshes. This implied that for different boundary layer thicknesses, different block structures would be necessary, adding complexity to the preprocessing procedure and reducing flexibility for general application.

The meshes illustrated in Fig. 5 were generated using GAMBIT's boundary layer mesh function, showing the independent alignment of boundary layer and main mesh. However, the feature did not work reliably and despite using the latest release, close discussions with the code vendors and intensive trials, the desired meshes could not be achieved in all cases. Correct definition of an independent alignment of boundary layer mesh and main mesh, often still resulted in the boundary layer running parallel to the main mesh. A significant effort was necessary to establish semi-reliable procedures in order to achieve the correct mesh structures for the most important cases. If a flawless mesh file for the CFD modules was available, scaling and mapping may be used to create similar face meshes. Care had to be taken when remeshing edges. Solving the problem was beyond the scope of this study and the code vendors were made aware of the problem. But for future activities involving structured meshes and boundary layers, other mesh generators should be considered.

CFD Module of the Intake. Initial trials were performed using a 1/4 circular shape of the computational domain, where when rotated around the symmetry axis, the intake faced into a semisphere. Due to a region of very sheared cells just upstream of the intake, a rectangular domain proved beneficial. The size of the domain was investigated in the direction upstream of the engine and perpendicular to the axis, selecting the height (i.e., radius) of the axisymmetric domain.

The domain length upstream of the engine face had a relatively small influence on the pressure recovery of the intake. Similar levels were predicted for domain lengths of eight times the fan diameter $D_{2,\mathrm{tip}}$ for different grid densities. The node density upstream was investigated for a wider range and a plateau of the pressure recovery factor was reached for densities in excess of 12 nodes/m.

The height of the domain was investigated in a similar manner. However, it was an aim that, once a CFD module was established, it could be applied to any operating point. The nature of the flow at the engine intake varies significantly between the extreme condition of static operation and top speed. The essential change is the mass flow ratio (MFR) and has been well described by Seddon and Goldsmith [12]:

- At ground running ($\mathrm{MFR}\rightarrow\infty$) the streamlines into the fan face may originate from upstream or downstream of the intake. In computational terms the flow might enter from the front, top, or rear of the domain. This raises issues for the definition of boundary types. Furthermore, the flow may not be able to follow the strong curvature around the intake lip and flow-separation zones may appear. This results in a significant reduction of the pressure recovery factor of the intake.
- At lower flight speeds during climb a higher power setting may lead to an operation in suction mode ($\mathrm{MFR}>1$).
- For high speed cruise the intake is generally designed such that the internal duct performance is at its optimum. The intake operates in spillage mode in a typical range of $\mathrm{MFR}=0.5\ldots0.8$. At such low mass flow ratios separation is

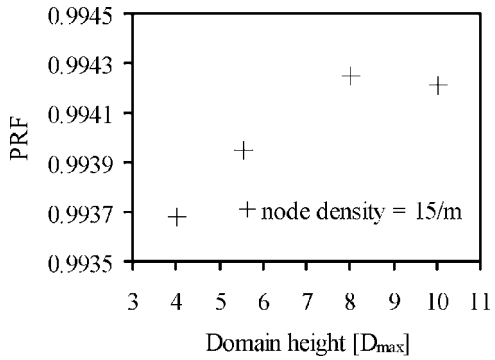


Fig. 6 Pressure recovery factor for different domain heights at static conditions

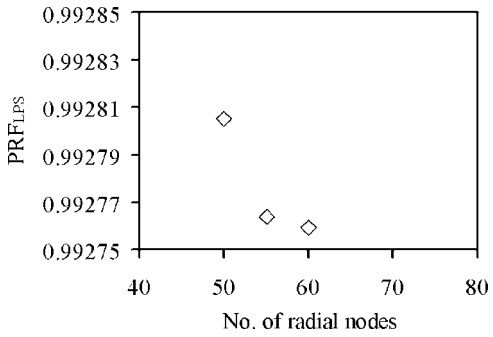


Fig. 7 Pressure recovery factor of the bypass-section versus radial nodes

avoided and the pressure recovery factor at its maximum within the flight envelope. All streamlines entering the fan face originate from the upstream boundary of the computational domain.

- Under top speed conditions, the mass flow ratio can reduce even further.

Generating a computational domain coping with the entire flight envelope, required that the domain height in particular was examined for static conditions. The pressure recovery factor versus the domain height for sea-level-static conditions is shown in Fig. 6. The height is referred to the maximum nacelle diameter D_{max}. For heights above $8D_{max}$ the pressure recovery factor was predicted to the fourth digit. A plateau appeared for node densities in excess of 20 nodes/m. However, the pressure recovery was predicted to the fourth digit with a deviation of 0.003% relative to the case of higher node density. This was considered to be close enough to the plateau values and a node density of 15 nodes/m was found to be sufficient for further cases.

The necessary domain height for conditions at altitude was significantly smaller with in the region of $4–5.5D_{max}$ in spillage mode. In order to save computational time for the investigated cases under cruise and end of climb conditions, using a domain height of $5.5D_{max}$ reduced the grid size by 26%. The boundary conditions were set as pressure conditions and were derived from the predicted values of the performance simulation. For the top of the domain, however, a symmetry condition which aligned the flow with the boundary proved favorable for engines at higher speeds. For the static conditions the larger domain was used and the boundary type at the top of the domain was changed to a pressure inlet condition in order to allow streamlines from this boundary to enter the engine inlet.

The density of the grid in the internal duct of the inlet was investigated for a range of 60–90 nodes along the span of the fan. A plateau of predicted pressure recovery factor was indicated for 90 or more nodes along the span. As a higher number of nodes led to higher aspect ratios of the cells far upstream, 90 nodes were considered to be sufficient. A prediction of the pressure recovery factor to the third digit was reached for 70 and more nodes. The prediction of the fourth digit was expected with 90 nodes along the span.

Turbulence was modelled for all the CFD modules using the Spalart–Allmaras model, which is implemented in FLUENT using wall functions. The turbulence intensity was set to 1%. This combination resulted in realistic profiles at the fan inlet, whereas higher turbulence intensities showed local pressure overshoots at the boundary layer edge. Other turbulence models showed exceedingly higher losses or comparable results at the price of higher computational effort.

All cases have been performed with the coupled implicit formulation of FLUENT. The discretisation schemes for the flow and turbulent viscosity were solved using the second order upwind scheme.

CFD Module of the Bypass. Similar considerations had to be taken into account to set up the numerical model for the bypass and nozzle. Using a rectangular domain for the study, the necessary length of the domain downstream of the engine needed to be established. As only the pressure loss within the low-pressure system was considered in this study, the dependency of this loss was investigated for different domain lengths. For the purpose of this study a domain length of around 16 D_{max} was considered to be sufficient, which enabled the prediction of the bypass pressure loss to the fourth digit.

It was essential to provide a mesh within the bypass duct that produced grid-independent predictions of the pressure loss. Therefore, the overall pressure loss from fan exit to the bypass-nozzle exit was calculated for different numbers of radial nodes along the bypass height. The results are shown in Fig. 7. The slope indicated that for 60 radial nodes, the pressure recovery for the low-pressure system was predicted to the third digit, possibly the fourth. A further increase of radial nodes would have required a further grid refinement in the streamwise direction to avoid high aspect ratios for the cells near the wall. This would have increased the mesh size significantly. A mesh with 60 radial nodes was therefore chosen for the study.

It has to be stated that the numerical model of the bypass did not include the outlet guide vanes of the fan and swirl was not included in the model. The predicted loss was therefore small. For future investigations, a porous pressure jump could be included at the location of the outlet guide vanes to model a pressure reduction at the bypass entry.

Matching of Performance and CFD Simulations

The idea of combining CFD modules and performance simulation required an iteration procedure. The CFD module supplied with ideal boundary condition, predicted the pressure loss for a certain mass flow. The performance code simulated the mass flow reduction due to the pressure loss after the engine rematched itself. Therefore, the pressure loss and mass flow had to be matched between the two tools and this was achieved by manual iteration.

As the performance tool provided only a total flow condition and the CFD modules required static conditions at the pressure exits, the static conditions were calculated using the Q function. For the first CFD simulation, the boundary conditions from the performance tool were taken directly. For further iterations, a flow discharge coefficient based on the last CFD iteration was included to estimate the required static pressure. Matching for the intake was usually achieved within ten iterations and a typical convergence plot of the matching procedure is shown in Fig. 8.

The residuals were calculated based on the deviation from the previous iteration step. The even numbers of iteration steps represented results from a CFD simulation, whereas the odd numbers

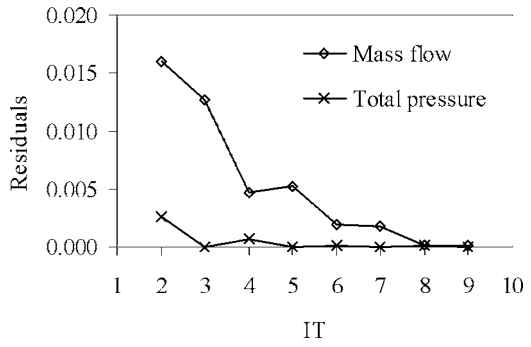


Fig. 8 Residuals for intake matching between performance and CFD tools

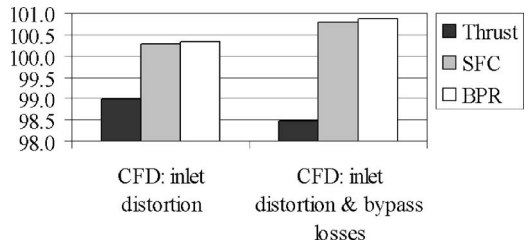


Fig. 9 Key whole-engine performance data for conventional 0D and 2D performance simulation using CFD modules

stood for performance predictions. It can be seen that the pressure residuals reach zero with every performance prediction, but the pressure loss had to be recalculated using the CFD module. When both residuals were close to zero, a matching was achieved. For the present study, the convergence criteria was set to 0.01% deviation in mass flow and 0.0% in total pressure.

The fan-exit profiles for the total pressure and total temperature were provided from the performance tool. This implied that for the bypass duct, the level of static pressure at the core inlet controlled the fan-exit mass flow and the core-inlet mass flow. The bypass nozzle was running at close to choked conditions and therefore the changes of bypass mass flow were small. The matching procedure was more complex and required slightly more, with up to 15 iterations (equivalent to seven CFD simulations).

The fan-exit profiles provided from the performance code were defined for 19 radial stations, whereas the grid-independent CFD module counted 90 nodes along the fan-exit span. To provide approximate data for each node, curve fitting was performed and $P_3(r)$ and $T_3(r)$ was derived for each CFD node. This was initially done in EXCEL spreadsheets, but later integrated into a MATLAB procedure to ease user application.

To take advantage of the losses provided by the CFD simulations for the bypass and bypass nozzle, the 0D bypass-nozzle component of the performance code was changed such that the discharge coefficient and pressure loss predicted from the CFD was used instead of the original characteristic. The losses of the bypass duct and bypass nozzle were combined, which required to derive the mass-averaged total pressure at the bypass inlet only. For the present study, this was evaluated using an EXCEL spreadsheet for the fan-exit total pressure distribution and the radial mass flow distribution (from the CFD simulation). Based on the bypass ratio predictions of the performance model, the coordinate of the split between the bypass and core flow was calculated and the mass-averaged total pressure for the bypass section was derived.

A perfect matching between the CFD modules and the performance code for the bypass duct was not achieved due to mass flow discrepancies between the radial fan performance model and the CFD simulations. The best match for cruise conditions showed mass flow mismatches of 0.54%, 0.15%, and 0.99% for the bypass duct, core inlet, and fan exit, respectively.

Performance Results

The engine performance results for cruise conditions are shown in Fig. 9. Including the CFD predictions of the inlet pressure loss of 0.44% into the model, resulted in a reduction of thrust by 1.02% and an increased SFC of 0.29%. The additional 2D modeling of the bypass loss resulted in a pressure loss of 0.43% which reduced the thrust by a further 0.49% and resulted in a further increase of SFC of 0.10%.

For cruise conditions, the intake loss is at a relatively low level. The resulting effect on overall engine performance was expected to be equivalently small. However, the results show that the cost-effect ratio is greater than one in terms of thrust degradation. Along with an increased SFC, the thermodynamic effect is therefore worth considering.

The radial fan-inlet profile of the matched solution is shown in Fig. 10 (left). The undistorted fan pressure ratio profile and the resulting profile when the CFD intake module and the engine performance model were matched are illustrated on the right hand side. As might be expected intuitively, the tip distortion had the largest effect on the redistribution of the pressure ratio.

The prediction of hub distortion caused by the spinner had a smaller effect on the pressure ratio of the fan. It should be stated, that the inclusion of the boundary layer resulting from the spinner may be questionable: In case that the radial profiles for the fan performance component originated from test data, these would include this effect already. The performance of a fan cannot be measured without a spinner present and the consideration of the intake boundary layer caused by the spinner would lead to a double booking of these losses. As notional data were used for the present case, hub, and tip distortion were considered. A simple solution to overcome the double booking when a CFD module for the intake was used, was to assume unity for the pressure recovery factor (PRF) in the hub boundary layer or to continue the peak pressure of the profile to the hub.

Discussion

This study demonstrated the feasibility of assisting gas turbine performance simulation with CFD methods to achieve a more realistic engine model. The initial effort that had to be invested to design the intake and bypass components, including the numerical investigations to assure that the CFD modules were numerically appropriate, was relatively high. However, it allowed the engine performance to be investigated with a more realistic, higher order of dimension and was therefore considered to be applicable over a wider operating range than conventional characteristics-based methods. The design procedure can be easily adapted for other high-bypass turbofans and for similar bypass ratios, a scaling of the generated CFD modules may be possible. The structure of the

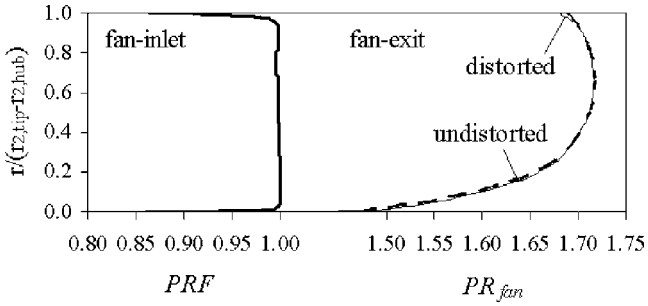


Fig. 10 Radial profiles of the intake pressure recovery and undistorted and distorted fan pressure ratio

modules allows for changes of the boundary layer mesh which should enable an adaptation to different flow conditions.

The CFD modules were a first step to a higher-fidelity model of the intake and bypass system, but were still based on major simplifications which limited the application to certain scenarios. Under the assumption of axisymmetry, only inlet flow parallel to the engine axis can be considered, allowing for radial distortion modeling. The more severe types of circumferential distortion, for example in crosswinds, cannot be modelled using this 2D approach.

In terms of the bypass, the CFD modules could not include the outlet guide vanes or the pylon. The predicted losses were therefore only due to friction losses in the "empty" duct and were significantly lower than in real engines. Also, swirl, which would normally enter the bypass from the fan, though compensated to a degree by the outlet guide vanes, has not been taken into account. A representative to the loss generated in the outlet guide vanes could be included into the model by using a so-called "porous pressure jump." This, in effect, calculates a pressure decrease introduced by a membrane of zero thickness as a function of permeability, thickness, and a pressure jump coefficient. In the absence of validation data, this appeared to be a crude approach and was therefore not considered in this study.

Even though this 2D approach for the CFD modules may appear very simple, the mesh for the intake and bypass system yielded in excess of 100 k cells. On a single processor, the first CFD calculation of an operating point usually required around 12 h. The boundary conditions of the far field had to be adjusted in several steps to the final values to avoid immediate divergence. This required user attention during the first 1–2 h. The first converged CFD solution was then taken as the initial solution for the next iteration between the performance code and the CFD simulations. These cases converged faster due to the closer proximity of the initial to the final solution. The next solution was achieved generally within 4–6 h. With a growing database of operating points for an investigated engine model, the convergence time of the CFD modules could be reduced and under best conditions, the time to reach a solution for a certain operating point may well be within one day.

Conclusions

- To overcome the lack of geometrical data, fundamental design principles were applied to create a generic, but representative intake and bypass geometry. The results for each component appeared reasonable and the design process can be adapted easily to other engines. The general purpose of using such data on this occasion was not related to achieving specific design improvements. A data set was required as a vehicle to demonstrate performance trends. In the academic environment, the lack of detailed design information is often an issue. But in similar cases, this simplified design procedure may help to overcome this hurdle.
- The CFD modules for the intake and bypass section were generated with considerable effort due to a software bug in the preprocessor of the CFD software. A meshing approach of structured grids with boundary layers used to provide multiblock structures can not be advised using GAMBIT. Alternatively, a different mesh generator or an unstructured approach may be used when starting a CFD simulation from scratch.
- The combination of performance simulation with CFD modules for the intake and bypass system has shown correct qualitative results. In the absence of information about pressure losses for these components, it is expected that valid results can be achieved for a wider operating range using the CFD modules.
- Radial inlet distortion affects the whole-engine performance. The thermodynamic effects may seem small, but show a greater than one cost-effect ratio in terms of thrust. Therefore, it appears well worth considering.
- The computational effort that had to be invested to achieve the results was quite high. A matched result between the performance code and the CFD modules required several days. This is most likely to be outside the time scopes in an industrial environment. However, previous experiences can be used to improve initial guesses and may speed up the procedure such that it could be performed within one day.

Acknowledgment

Thanks go to Darrell D. Williams for his kind advice regarding the design of the intake.

Nomenclature

a	=	stream-wise distance from highlight to throat
b	=	$r_H - r_{\text{th}}$
BLE	=	boundary layer edge
BPR	=	by pass ratio
CFD	=	computational fluid dynamics
CR	=	contraction ratio
D	=	diameter
DR	=	diffusion ratio
MFR	=	mass flow ratio
NPSS	=	numerical propulsion system simulation
P	=	total pressure
PR	=	pressure ratio
PRF	=	pressure recovery factor
r	=	radius
SFC	=	specific fuel consumption
T	=	total temperature
TET	=	turbine entry temperature
β	=	boat-tail angle

Subscripts

1…10	=	station numbers
a	=	afterbody
f	=	forebody
gr	=	gross
H	=	highlight
LPS	=	low-pressure system
th	=	throat

References

[1] Hall, E. J., Delaney, R. A., Lynn, S. R., and Veres, J. P., 1998, "Energy Efficient Engine Low Pressure Subsystem Aerodynamic Analysis," Technical Memorandum TM-1998-208402, NASA.

[2] Riegler, C., Bauer, M., and Kurzke, J., 2001, "Some Aspects of Modeling Compressor Behavior in Gas Turbine Performance Calculations," Very Large Data Bases, **123**(2), pp. 372–378.

[3] Curnock, B., Yin, J., Hales, R., and Pilidis, P., 2001, "High-Bypass Turbofan Model Using a Fan Radial-Profile Performance Map," Aircraft Design, **4**(2–3), pp. 115–126.

[4] Marshall, C. D., 1998, "The Modelling of High Bypass Turbofans," Master's thesis, Cranfield University, UK.

[5] Yin, J., and Pilidis, P., 2002, "Influence of Inlet Profile on High-BPR Turbofan Performance Using a Radial Profile Map," *Proceedings of the 23rd International Council of Aerospace Sciences*, Toronto, Canada, Sept. 6–13 2002, ICAS 2002-R79.

[6] Li, M. S., 2004, "2D Low Bypass-Ratio Turbofan Modelling," Ph.D. thesis, Cranfield University, UK.

[7] Cumpsty, N., 2000, *Jet Propulsion*, Cambridge University Press, Cambridge, UK.

[8] RTO, 2002, "Performance Prediction and Simulation of Gas Turbine Engine Operation," RTO Technical Report 044, NATO Research and Technology Organisation, Neuilly-Sur-Seine Cedex, France.

[9] McKenzie, A. B., 1997, *Axial Flow Fans and Compressors*, Cranfield Series on Turbomachinery Technology, Ashgate Publishing, Aldershot, UK.

[10] Albers, J. A., and Miller, B. A., 1973. Effect of Subsonic Inlet Lip Geometry on Predicted Surface and Flow Mach Number Distributions, Technical note NASA TN D-7446, NASA.
[11] ESDU, 1981, Drag of Axisymmetric Cowls at Zero Incidence for Subsonic Mach Numbers, Data item 81024, ESDU International, London, UK.
[12] Seddon, J., and Goldsmith, E. L., 1999, *Intake Aerodynamics*, 2nd. ed., Blackwell Science, Oxford, UK.
[13] Stanhope, F. W., 1968, "The Performance of NACA 1-Series Intakes," Power plant research report IAR 85002, Rolls Royce, UK.
[14] ESDU, 1994, NACA 1-Series Geometry Representation for Computational Fluid Dynamics, Data item 94013, ESDU International, London, UK.
[15] Williams, D., 2004, "Propulsion Systems Performance and Integration," "Thermal Power," Cranfield University, Lecture Series Cranfield, UK.
[16] Walsh, P. P., and Fletcher, P., 1999, *Gas Turbine Performance*, Blackwell Science, Oxford, UK.
[17] FLUENT, 2003, FLUENT 6.0 Documentation, FLUENT Inc.

Physical Interpretation of Flow and Heat Transfer in Preswirl Systems

Paul Lewis
e-mail: p.r.lewis@bath.ac.uk

Mike Wilson
e-mail: m.wilson@bath.ac.uk

Gary Lock
e-mail: g.d.lock@bath.ac.uk

J. Michael Owen
e-mail: j.m.owen@bath.ac.uk

University of Bath,
Bath BA2 7AY, UK

This paper compares heat transfer measurements from a preswirl rotor–stator experiment with three-dimensional (3D) steady-state results from a commercial computational fluid dynamics (CFD) code. The measured distribution of Nusselt number on the rotor surface was obtained from a scaled model of a gas turbine rotor–stator system, where the flow structure is representative of that found in an engine. Computations were carried out using a coupled multigrid Reynolds-averaged Navier-Stokes (RANS) solver with a high Reynolds number k-ε/k-ω turbulence model. Previous work has identified three parameters governing heat transfer: rotational Reynolds number (Re_ϕ)*, preswirl ratio* (β_p)*, and the turbulent flow parameter* (λ_T)*. For this study rotational Reynolds numbers are in the range* $0.8\times10^6<\mathrm{Re}_\phi<1.2\times10^6$*. The turbulent flow parameter and preswirl ratios varied between* $0.12<\lambda_T<0.38$ *and* $0.5<\beta_p<1.5$*, which are comparable to values that occur in industrial gas turbines. Two performance parameters have been calculated: the adiabatic effectiveness for the system,* $\Theta_{b,\mathrm{ad}}$*, and the discharge coefficient for the receiver holes,* C_D*. The computations show that, although* $\Theta_{b,\mathrm{ad}}$ *increases monotonically as* β_p *increases, there is a critical value of* β_p *at which* C_D *is a maximum. At high coolant flow rates, computations have predicted peaks in heat transfer at the radius of the preswirl nozzles. These were discovered during earlier experiments and are associated with the impingement of the preswirl flow on the rotor disk. At lower flow rates, the heat transfer is controlled by boundary-layer effects. The Nusselt number on the rotating disk increases as either* Re_ϕ *or* λ_T *increases, and is axisymmetric except in the region of the receiver holes, where significant two-dimensional variations are observed. The computed velocity field is used to explain the heat transfer distributions observed in the experiments. The regions of peak heat transfer around the receiver holes are a consequence of the route taken by the flow. Two routes have been identified: "direct," whereby flow forms a stream tube between the inlet and outlet; and "indirect," whereby flow mixes with the rotating core of fluid.* [DOI: 10.1115/1.2436572]

1 Introduction

The blade-cooling air in gas turbines is usually supplied to the rotating high-pressure blades by stationary preswirl nozzles. The cooling air is swirled, which reduces the work done by the rotating turbine disk in accelerating the air to the disk speed. This in turn reduces the total temperature of the air entering the receiver holes in the disk. A simplified diagram of the so-called direct transfer preswirl system is shown in Fig. 1.

The designer is interested in calculating the pressure drop and cooling effectiveness of the preswirl system. There is also a need to understand the heat transfer between the cooling air and the turbine disk, particularly the possible creation of local nonuniform temperatures in the metal that could lead to large thermal stresses.

Meierhofer and Franklin [1], who were the first to measure the effect of preswirl on the temperature drop in a direct-transfer system, showed that swirling the air could significantly reduce the total temperature in the receiver holes of a turbine disk. El-Oun and Owen [2] developed a theoretical model for the so-called adiabatic effectiveness, $\Theta_{b,\mathrm{ad}}$, based on the Reynolds analogy. The model, which was in good agreement with the temperatures measured on their rotating-disk rig, showed that $T_{t,b}$, the total temperature in the receiver holes, decreased monotonically as β_p, the preswirl ratio, increased even when β_p was significantly greater than unity.

Geis et al. [3] made measurements of the adiabatic effectiveness, which showed that the measured values of $T_{t,b}$ were significantly higher than the values predicted from their ideal model. (It should be pointed out that their preswirl ratio was based on isentropic values rather than on measurements.) Chew et al. [4] made numerical simulations of both the "Karlsruhe rig," used by Geis et al., and a "Sussex preswirl rig." The computations were in good agreement with the results of both rigs, and the low adiabatic effectiveness of the Karlsruhe rig was attributed to the geometry of the preswirl chamber; in particular, the Karlsruhe rig had a much larger stator area, which reduced the effective swirl ratio and consequently reduced the effectiveness.

Chew et al. [4] and Farzaneh-Gord et al. [5] independently derived theoretical models for the adiabatic effectiveness of a direct-transfer system, taking account of the moment on the stator. (These models predict lower values of $\Theta_{b,\mathrm{ad}}$ than that of Karabay et al. [6], who based their model on a cover-plate system in which the preswirl air flows radially outward between two rotating disks.)

Popp et al. [7] carried out a CFD analysis of a cover-plate system, computing the temperature drop and the discharge coefficients for different geometries. They showed that C_D, the discharge coefficient for the receiver holes, became a maximum when the relative tangential velocity was close to zero. This effect

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received July 19, 2006; final manuscript received July 20, 2006. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2006: Land, Sea and Air (GT2006), Barcelona, Spain, May 8–11, 2006. Paper No. GT2006-90132.

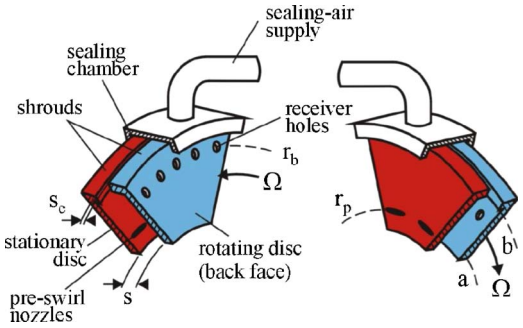


Fig. 1 Schematic diagram of test section

was confirmed experimentally by Dittmann et al. [8] who were the first to measure the discharge coefficients in a direct-transfer system.

Yan et al. [9] measured the discharge coefficients for the receiver holes of a direct-transfer system for a range of rotational speeds and flow rates. For $\beta_1 < 1$ (where β_1 is the measured swirl ratio upstream of the receiver holes) C_D increased monotonically as β_1 increased from $\beta_1 \approx 0.3$ to 0.9. They also found, as did Popp et al., that C_D depends on the ratio of the area of the receiver holes to that of the nozzles; for a given value of the preswirl ratio, β_p, C_D increases as the area ratio decreases. (It should be noted that, owing to a printer's error, the wrong figures were printed in Ref. [9]; the correct figures are given in Ref. [10]).

Heat transfer in a direct transfer rig was studied experimentally and computationally by Wilson et al. [11] using fluxmeters to determine the local Nusselt numbers. Their axisymmetric CFD results gave reasonable predictions of the velocity and temperature in the core but underpredicted the measured Nusselt numbers.

Numerous experimenters have used thermochromic liquid crystal (TLC) to determine heat transfer coefficients on purpose-built test sections. A common technique is to solve Fourier's transient conduction equation to calculate h for a semi-infinite solid exposed to a step change in air temperature. As it is virtually impossible to achieve a step change in the air temperature of preswirl rigs, Newton et al. [12] developed the so-called "slow transient" technique. Lock et al. [13,10] used this technique to measure the local Nusselt numbers, on the rotating disk of a direct-transfer rig, for a range of rotational speeds, flow rates, and preswirl ratios. The measurements showed that Nu was virtually axisymmetric except near the receiver holes, where large variations occurred. They also found that there were two flow regimes: at the larger preswirl flow rates, inertial effects dominated and the flow impinged on the rotating disk creating a peak in Nu; at the smaller flow rates, viscous effects dominated and boundary-layer flow controlled the heat transfer.

In this paper, a commercial three-dimensional (3D) CFD code is used to compute the flow and heat transfer in the "Bath rig" used by Lock et al. The computations are compared with measured values of the discharge coefficients, with theoretical values of adiabatic effectiveness and with measured local Nusselt numbers. In particular, the computations are used to give a physical insight into the complex flow and heat transfer that occurs in these direct-transfer preswirl systems.

2 Governing Parameters

Owen and Rogers [14] showed that, for a rotating cavity, the turbulent flow structure depends on only two nondimensional parameters: the inlet swirl ratio, β_p, and the turbulent flow parameter, λ_T. These are defined as

$$\beta_p = \frac{v_{\phi,p}}{\Omega r_p} \tag{1}$$

and

$$\lambda_T = c_w \, \mathrm{Re}_\phi^{-0.8} \tag{2}$$

where

$$c_w = \frac{\dot{m}_p}{\mu b} \tag{3}$$

and

$$\mathrm{Re}_\phi = \frac{\rho \Omega b^2}{\mu} \tag{4}$$

The value $\lambda_T = 0.22$ corresponds to the flow rate entrained by a disk rotating in an infinite environment, the so-called free disk. For turbine-blade cooling systems, $\lambda_T \approx 0.4$ and $\beta_p \approx 1$.

For the Bath rig (described in Sec. 3) it follows that

$$\beta_p = \frac{C}{N} \frac{c_w}{\mathrm{Re}_\phi} = \frac{C}{N} \frac{\lambda_T}{\mathrm{Re}_\phi^{0.2}} \tag{5}$$

where C is a geometric constant given by

$$C = \frac{4b^3 \cos\theta}{\pi d^2 r_p} \tag{6}$$

It can be seen that the preswirl ratio is not independently variable: it depends upon N, the number of preswirl nozzles, and upon λ_T and Re_ϕ. In the experiments of Yan et al. [9], two values of N were used: $N = 12$ and 24. In the experiments discussed below, $N = 24$.

The Bath rig uses a simplified engine geometry, and tests were conducted at representative values of β_p and λ_T, thereby producing flow structures typical of those found in engines. However, in engines Re_ϕ is on the order of 10^7, which is an order of magnitude greater than that achievable in the rig. As the heat transfer depends strongly on Re_ϕ, as well as on β_p and λ_T, the rig Nusselt numbers will be much smaller than those found in engines. This is discussed further in Sec. 6.

3 Experimental Method

Experimental results were produced by Yan et al. [9] and Lock et al. [13,10] using the "slow transient" TLC technique described by Newton et al. [12]. The salient points of the experimental method are presented here for convenience and a section of the experimental rig is illustrated in Fig. 1.

The rotor is a transparent polycarbonate disk with a radius of 0.216 m, allowing optical access to the wheel space. The disk has 60 circular receiver holes with centers at a disk radius of 0.200 m. To reduce heat transfer from the air inside the receiver holes, the holes are filled with Rohacell (low-conductivity foam) bushes producing an effective receiver hole diameter of 8.0 mm. The disk has a thickness of 10 mm, and the receiver holes, which have a length to diameter ratio of 1.25, vent directly into the laboratory. A shroud of carbon fiber surrounds the rim of the disk and rotates with it. A section of the rotor surface inside the wheelspace is painted with thermochromic liquid crystal.

The stator is also a polycarbonate disk, which is mounted onto an aluminum disk. The gap between the rotor and stator is 11 mm ($G = 0.051$), and the clearance between the rotating and stationary shrouds is 1 mm. The air pressure in the wheel space is balanced by sealing air to restrict leakage or ingress. The preswirl nozzles comprise 24 circular holes, of 7.1 mm diameter, drilled at an angle of 20 deg to the tangential direction and at a radius of 160 mm. A stationary Rohacell hub forms the inner boundary of the wheel space at a radius of 0.145 m.

The radial variation of pressure and tangential velocity is measured by a combination of pitot tubes, located at nine radial stations in the midplane ($z/s = 0.5$), and static pressure tappings at the same radii on the stator. A total-temperature probe and pitot tube are also located in a nozzle outlet to measure the temperature and velocity of the inlet flow. More details can be found in Yan et al. [9].

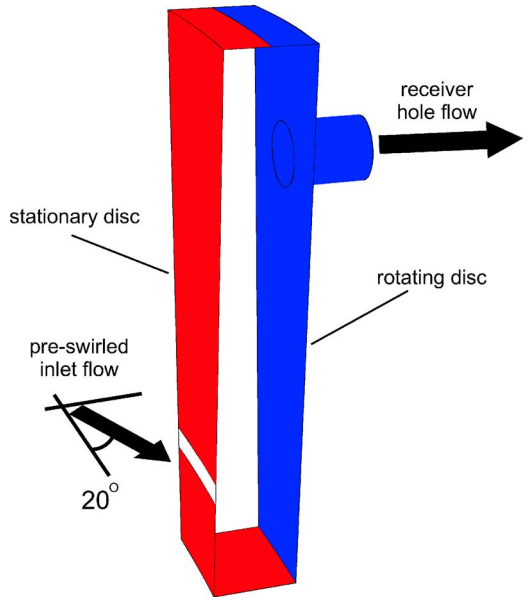


Fig. 2 Schematic diagram of computational domain

The main air supply to the system is passed through a mesh heating element which essentially creates a step change in temperature immediately downstream of the heater; owing to heat losses in the ducting, the temperature rise of the air downstream of the preswirl nozzles is exponential. A strobe light is used to illuminate the rotor and the resulting transient disk temperature distribution is captured on digital video at 25 frames/s. The red–green–blue (RGB) signals of each frame are converted to hue and analyzed to calculate the temperature history and steady-state heat transfer coefficient, as described by Newton et al. [12].

4 Computational Method

The computational domain is designed to be a realistic representation of the experimental rig described above, the only exception being the preswirl nozzles. To enable steady-state computations, the preswirl nozzles are modeled as an axisymmetric slot having the same inlet area, and therefore the same inlet velocity, as the nozzles. The model contains a 1/60th sector of the experimental rig, i.e., a 6 deg section enclosing one of the receiver holes. Boundaries have either a no-slip condition or a periodic interface applied, and walls are defined as rotating or stationary as required. Clearances at no-flow boundaries were taken to be zero.

Figure 2 shows the computational domain, the red shaded areas being stationary while the blue shaded areas rotate with angular velocity Ω. Axial and circumferential velocities are prescribed at the inlet to give the flow angle of 20 deg to the tangential direction. Velocity and static temperature values were prescribed to match measurements made at the nozzle. At the outlet a static pressure boundary condition was used.

The commercial code used for this investigation is CFX5.7, a finite volume, coupled algebraic multigrid solver. The advection scheme is second-order accurate based on the method of Barth and Jesperson [15]. The energy equation is solved, including the viscous work term, and variable density effects taken into account. Buoyancy effects within the wheel space are ignored.

The mesh is a hybrid of unstructured tetrahedral elements, with prismatic elements near the wall. Delaunay triangulation is used to create the surface mesh followed by an advancing front volume mesher. A mesh sensitivity study was carried out to confirm that the fluid dynamics and heat transfer results are not grid dependent. Typical y^+ distributions, for computations both in the viscous and inertial regime, are shown in Fig. 3, and are within the required range for the turbulence model.

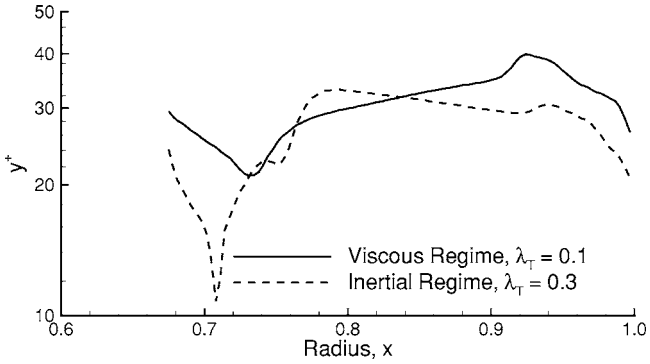


Fig. 3 Typical radial distribution of y^+ on the rotor for inertial and viscous regime computations

The turbulence model used is the high-Reynolds number baseline (BSL) model of Menter [16]. This is a blend of a k-ω formulation with wall functions (Ref. [17]), in the near wall region, and a k-ε model away from the wall. This overcomes sensitivities to freestream turbulence levels normally experienced by k-ω models [18]. The convective heat transfer model applied to the rotor wall is based on the method of Kader [19]; and other surfaces were assumed adiabatic.

Additional computations were performed to confirm that the results were not sensitive to inlet turbulence level or to the uniform rotor temperature boundary condition used to approximate the time varying distribution that occurred in the experiment.

5 Fluid Dynamics

5.1 Velocity and Pressure. Figure 4(a) shows a comparison between the computed and measured radial variation of β_∞, the nondimensional swirl ratio at the midplane ($z/s=0.5$). The maximum value of β_∞, which occurs at the inlet radius due to the flow from the nozzles, is well predicted by the computations. However, at the larger radii, the computations overpredict the measured values. This may be due in part to overprediction of turbulence levels by the high-Reynolds number turbulence model.

Figure 4(b) shows good agreement between the computed and measured radial distribution of static pressure. In the rotating core of fluid, away from the rotor and stator, $dp/dr=\rho v_\phi^2/r$, and as a consequence the static pressure increases radially. The overprediction of the total pressure in Fig. 4(c) is caused mainly by the overprediction of β_∞ referred to above.

The agreement between computations and measurements shown in Fig. 4(a) is best for the lowest value of λ_T shown, for which the flow is in the viscous regime. The computed mixing at the higher values of λ_T (for which the flow is in the inertial regime) may be affected by the use of a high-Reynolds-number turbulence model and the simplified slot geometry at inlet. (Yan et al. [9], who used a discrete nozzle inlet and a low-Reynolds-number k-ε turbulence model, obtained better agreement with measurements than that shown in Fig. 4(a).)

5.2 Adiabatic Effectiveness. The adiabatic effectiveness, $\Theta_{b,\mathrm{ad}}$, is defined as

$$\Theta_{b,\mathrm{ad}} = \frac{c_p(T_{0,p} - T_{t,b})}{0.5\Omega^2 r_b^2} \tag{7}$$

For given inlet conditions the total temperature of the air in the rotating receiver holes, $T_{t,b}$, decreases linearly as $\Theta_{b,\mathrm{ad}}$ increases.

Karabay et al. [6] derived a theoretical value for $\Theta_{b,\mathrm{ad}}$ using the first law of thermodynamics; the work done on, or by, the air was proportional to the moment required to change the tangential velocity of the air from $\beta_p\Omega r_p$, at the preswirl nozzles, to Ωr_b in the receiver holes. Their equation, which was derived for a *cover-*

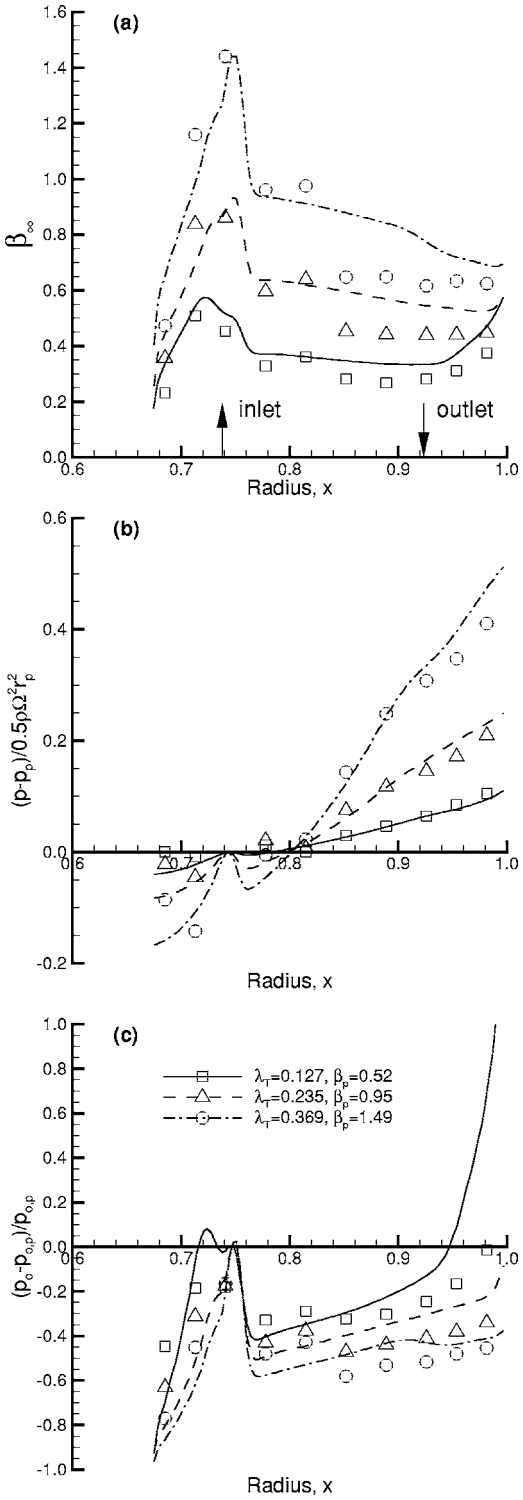


Fig. 4 Comparison of computed (lines) and measured (symbols) results for swirl ratio and pressure: $Re_\phi = 0.8\times10^6$

plate system in which there is no stator to reduce the swirl, is given by

$$\Theta_{b,\mathrm{ad}} = 2\beta_p\left(\frac{r_p}{r_b}\right)^2 - 1 \tag{8}$$

It should be noted that $\Theta_{b,\mathrm{ad}} \propto \beta_p$: $\Theta_{b,\mathrm{ad}}$ increases, and $T_{t,b}$ decreases, monotonically with β_p even when $\beta_p > 1$. Karabay et al. also used CFD to compute $\Theta_{b,\mathrm{ad}}$ in a cover-plate system for $0 \leq \beta_p \leq 3$, and their computations were in good agreement with this equation.

Farzaneh-Gord et al. [5] modified Eq. (8) for a *direct-transfer system* to account for the moment on the stator, M_s. Their model gives

$$\Theta_{b,\mathrm{ad}} = 2\beta_p\left(\frac{r_p}{r_b}\right)^2 - 1 - \frac{M_s}{1/2\dot{m}\Omega r_b^2} \tag{9}$$

where the effect of M_s (a positive quantity) is to reduce $\Theta_{b,\mathrm{ad}}$.

Figure 5(*a*) shows a comparison between the computed and theoretical values of $\Theta_{b,\mathrm{ad}}$ for the present case, where $r_p/r_b = 0.8$. The computations based on Eq. (7) use the bulk-average relative total temperature computed at the outlet from the receiver holes. These computations are in excellent agreement with the theoretical values of $\Theta_{b,\mathrm{ad}}$ based on Eq. (9) for the direct-transfer system; the values of M_S in this equation were obtained by computing the tangential shear stress over all of the stationary surfaces in the rig geometry. The theoretical values of $\Theta_{b,\mathrm{ad}}$ calculated using Eq. (8) for a cover-plate system are significantly higher than those for the direct-transfer system.

The theoretical values of $\Theta_{b,\mathrm{ad}}$ are based on the assumption that the air achieves solid-body rotation in the receiver holes. Figure 5(*b*) shows that β_2, the computed bulk-average value of β at the outlet from the holes, is indeed close to unity. (The fact that the computed values of $\beta_2 \approx 1$ could be due to the assumed constant pressure condition at outlet from the receiver holes. In an engine, where the blade-cooling air flows radially outward from the receiver holes, solid-body rotation would be expected to occur; for a rig with short receiver holes, it is possible that solid-body rotation would not be achieved.)

It should be pointed out that both the computed and theoretical values of $\Theta_{b,\mathrm{ad}}$ for the direct-transfer rig depend, explicitly or implicitly, on the computed value of M_s. It is, therefore, uncertain that the theory will agree with experimental measurements of $\Theta_{b,\mathrm{ad}}$, which were not made in the tests reported here. The measurement of $\Theta_{b,\mathrm{ad}}$ is nontrivial: the actual (not the isentropic) value of β_p must be known; the bulk-average relative velocity and total temperature inside the receiver holes must be measured accurately; and the rotating disk should be made from a thermal insulator to reduce the heat transfer to the cooling air. These three conditions are seldom, if ever, achieved in rotating-disk rigs.

5.3 Discharge Coefficients. So as to be consistent with other research workers, the discharge coefficient for the receiver holes is defined as

$$C_D = \frac{\dot{m}_b}{\dot{m}_i} \tag{10}$$

where $\dot{m}_i$ is the isentropic mass flow rate. This can be calculated from the first law of thermodynamics, for an adiabatic system, taking into account the rate of work done by or on the air from Stations 1–2 in a stream tube. The equation derived by Yan et al. [9] is

$$\frac{\dot{m}_i}{A_2} = \rho_{0,1}\left(\frac{p_2}{p_{0,1}}\right)^{1/\gamma}\left\{\left(\frac{2\gamma}{\gamma-1}\right)\frac{p_{0,1}}{\rho_{0,1}}\left[1-\left(\frac{p_2}{p_{0,1}}\right)^{\gamma-1/\gamma}\right] + 2\Omega(r_2V_{\phi,2} - r_1V_{\phi,1}) - V_{\phi,2}^2\right\}^{1/2} \tag{11}$$

The first term inside the curly brackets is the standard result for compressible flow in a stationary nozzle; the second term is the work term resulting from the change of angular momentum of the air; the last term is due to the fact that the air in the receiver holes has an absolute tangential, as well as an axial, component of velocity. In their measurements, Yan et al. based $p_{0,1}$ and $V_{\phi,1}$ on their pitot-tube measurements at $r = r_b$ and $z/s = 0.5$. They took p_2

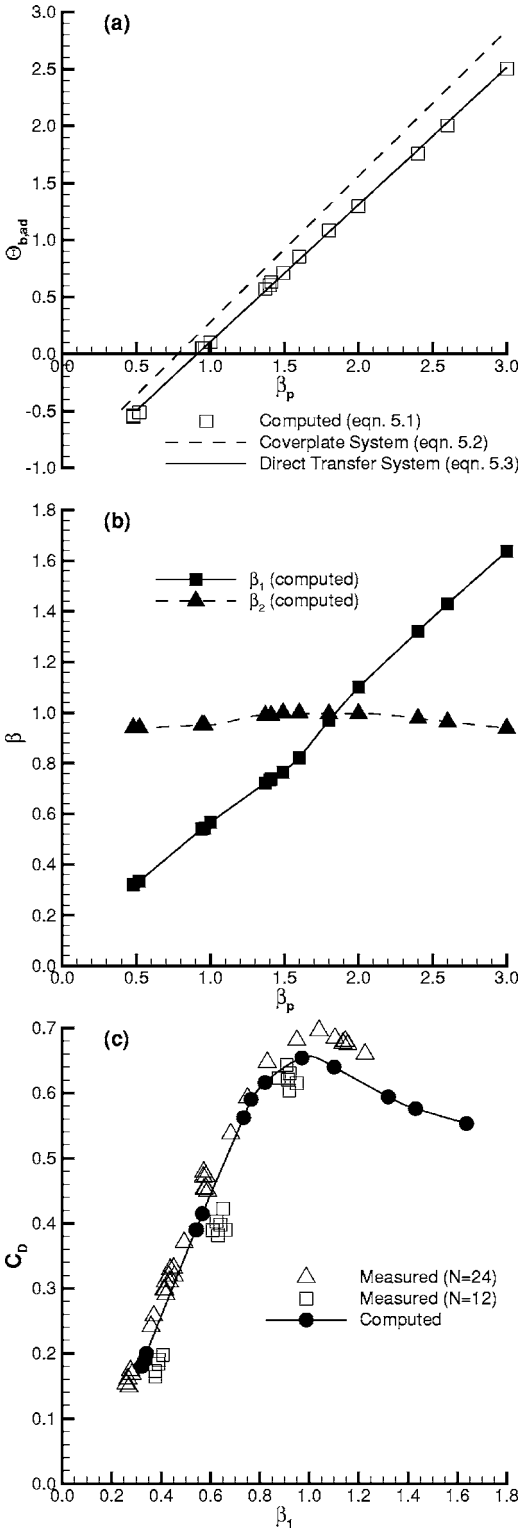


Fig. 5 Variation of $\Theta_{b,ad}$, β, and C_D with β_p for $0.8\times10^6<\mathrm{Re}_\phi<1.2\times10^6$ and $0.12<\lambda_T<0.38$: (*a*) comparison between computed and theoretical $\Theta_{b,ad}$; (*b*) comparison between computed β_1 and β_2; and (*c*) comparison between computed and measured C_D

as the atmospheric pressure at outlet from the receiver holes, and assumed that $V_{\phi,2}=\Omega r_b$. The same assumptions and locations were used here in determining the computed values of $\dot{m}_i$.

As shown in Fig. 5(*b*) the computed values of β_2 are close to unity, supporting the assumption of solid-body rotation at the outlet from the receiver holes. The computed values of β_1, which are based on the bulk-average values of $V_{\phi,1}$ calculated at $r=r_1=r_b$ and $z/s=0.5$, are significantly smaller than β_p, and it should be noted that $\beta_1=1$ when $\beta_p\approx1.8$.

Figure 5(*c*) shows a comparison between the computed and measured variation of C_D with β_1. The measured values were obtained with 24 and 12 preswirl nozzles ($N=24$ and $N=12$); the $N=12$ results were obtained by blocking alternate nozzles. As for all computations presented in this paper, the computed results were obtained for the case where the annular slot in the stator had an area equivalent to the $N=24$ tests.

As expected, the computations and experiments show that C_D reaches a maximum value at $\beta_1=1$; the computed maximum of $C_D=0.65$ is lower than that in the experiments, $C_D=0.70$. Dittman et al. [20] measured discharge coefficients for a rotating short orifice, with the same length to diameter ratio as the receiver holes, and found that C_D had a maximum value of 0.78 when the fluid and the orifice had the same circumferential velocity.

The maximum value of C_D occurs at a critical value of β_p, which will depend strongly on the system geometry. For the system used here, the critical value of β_p was approximately 1.8 for the computations and 2.3 for the experiments. The important practical consequence is that, although the cooling effectiveness increases monotonically as β_p increases, the flow rate of the blade cooling air will decrease if the critical value of β_p is exceeded.

6 Heat Transfer

6.1 Radial Variation of Nusselt Number. The computed heat flux is nondimensionalised to form the local Nusselt number, Eq. (13), based on the adiabatic disc temperature derived by Karabay et al. [6] for a cover-plate system and which is shown in Eq. (12)

$$T_{w,\mathrm{ad}}=T_{0,p}-\frac{V_{\phi,\infty}^2}{2C_p}+R\frac{\Omega^2r^2}{2C_p}\left(1-\frac{V_{\phi,\infty}}{\Omega r}\right)^2 \tag{12}$$

$$\mathrm{Nu}=\frac{q_w r}{k(T_w-T_{w,\mathrm{ad}})} \tag{13}$$

Equation (12) is a theoretical value based on the Reynolds analogy, and the values of $T_{w,\mathrm{ad}}$ computed by Karabay et al. [6] were in excellent agreement with this equation. Newton et al. [12] used wide-band TLC to measure the adiabatic-disk temperature on the Bath rig. Apart from the region near the preswirl nozzles, the differences between the measured and theoretical values of $T_{w,\mathrm{ad}}$ were mostly less than 0.5°C. Owing to the uncertainty in the wide-band TLC measurements, Eq. (12) was used to calculate $T_{w,\mathrm{ad}}$ for the measured Nusselt numbers.

Owen and Rogers [21] showed that, for turbulent boundary-layer flow in rotor-stator systems, $\mathrm{Nu}\propto\mathrm{Re}_\phi^{0.8}$. If the heat transfer in the Bath rig is controlled by turbulent boundary-layer flow then the parameter $\mathrm{Nu}\,\mathrm{Re}_\phi^{-0.8}$ would be expected to be independent of Re_ϕ.

Figure 6 shows a comparison between computed and experimental values of $\mathrm{Nu}\,\mathrm{Re}_\phi^{-0.8}$ for three values of λ_T and three values of Re_ϕ (As Eq. (5) shows, $\beta_p\propto\lambda_T$). The measured local Nusselt numbers were obtained by Lock et al. [13], and those shown in Fig. 6 were evaluated along a radial line midway between two adjacent receiver holes.

Figure 6(*a*), for $\lambda_T=0.13$ and $\beta_p=0.5$, shows a distinct peak in heat transfer near the receiver holes ($x=0.93$) but there is no sign of impingement near the preswirl nozzles ($x=0.74$). Lock et al. defined this as the viscous regime, and it can be seen that $\mathrm{Nu}\,\mathrm{Re}_\phi^{-0.8}$ is a good correlating parameter for both the computational and experimental results. Although the radial variation of the computations and measurements are qualitatively similar, the compu-

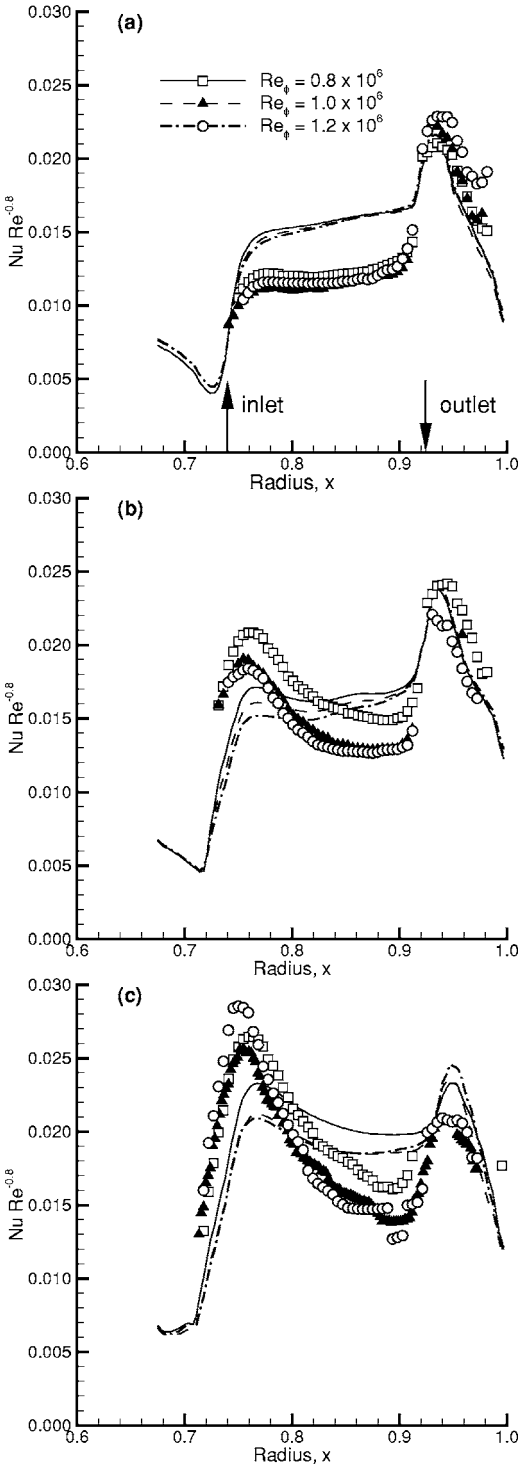


Fig. 6 Radial variation of Nu $Re_\phi^{-0.8}$: (*a*) λ_T=0.12, β_p=0.5; (*b*) λ_T=0.24, β_p=1.0; and (*c*) λ_T=0.35, β_p=1.5

tations overpredict the measured values except in the region near the receiver holes.

In Fig. 6(*b*) for $\lambda_T \approx 0.24$ and $\beta_p \approx 1.0$, the experimental results show a distinct peak near the nozzles, signifying that the flow is in the inertial regime. The computations show only a small peak at this radius. For these conditions, the parameter Nu $Re_\phi^{-0.8}$ fails to collapse either the experimental or computational data.

In Fig. 6(*c*) for $\lambda_T \approx 0.35$ and $\beta_p \approx 1.5$, both sets of results exhibit the inertial peak near the preswirl nozzles. The radial variation of the computations and measurements are qualitatively similar, but again the parameter Nu $Re_\phi^{-0.8}$ fails to collapse the data.

6.2 Circumferential Variation of Nusselt Number. Figure 7 shows comparisons of Nusselt number contours across an 18 deg sector of the rotor, as studied experimentally. Results on the right were produced by computation and those on the left are experimental results. The conditions for case 7(*a*) classify it within the viscous regime; the two other examples, 7(*b*) and 7(*c*), relate to the inertial regime.

Comparing Fig. 7(*a*) with the results in Fig. 6, the same difference in Nusselt number magnitude is visible and the improved agreement at the receiver hole radius is also apparent. In the region close to the edge of the holes there is an absence of experimental data caused by the presence of the Rohacell bushes described in Sec. 3. (The opaque bushes also cause shadows over the transparent rotor, obscuring the results in this region.)

In Fig. 7(*b*) and 7(*c*), representing the inertial regime, there is good qualitative agreement between computations and measurements at high radii.

A small region of high heat transfer is observable around the receiver holes in each case in Fig. 7. At low λ_T and β_p this region is located at the "9 o'clock position" with respect to the holes, Fig. 7(*a*). As λ_T and β_p are increased the region moves around towards the "11 o'clock position," Fig. 7(*c*). Luo et al. [22] performed temperature measurements around a rotating disk with receiver holes and observed similar behavior around the holes. (In an engine, high heat transfer in this region could result in thermal stresses within the rotor.)

6.3 Physical Interpretation of Heat Transfer Results. Using an axisymmetric CFD code, Wilson et al. [11] showed that air entered the receiver holes by "direct" and "indirect" routes. The former refers to flow traveling directly along a streamline connecting the inlet and the outlet, and therefore not mixing with the core flow. Indirect flow mixes with the core flow before entering the receiver holes.

This idea can be extended to the study of nonaxisymmetric systems by considering that the direct flow travels in a stream tube between the preswirl nozzles and the receiver holes. Computing streamlines for the direct flow allows the path of the stream tube to be evaluated. Figure 8 shows the stream tube relative to the rotor for a variety of conditions. The inner location is at the radius of the preswirl nozzles, and the outer location is at the radius of the receiver holes. These results show only a weak effect of Re_ϕ.

Figure 9 shows measured heat transfer results combined with streamlines calculated using the *full* velocity field from the computations. Figure 9(*a*) is a radial section with the preswirl inlet and stator on the left and the receiver hole and rotor on the right. The orange streamline shows that flow from the nozzle can be either direct or indirect: the direct flow exits through the receiver hole; the indirect flow continues to a higher radius and will recirculate in the system and mix with the core flow. The black streamline shows that indirect flow, which has entered the core, can either exit through the receiver hole or continue circulating in the core.

Figure 9(*b*) shows the same streamlines in a circumferential section. It can be seen that the flow at the rotor surface that is aligned with the receiver holes becomes direct flow. The remaining flow follows the indirect route.

Figure 9(*c*) shows an isometric view of the same streamlines. It is the flow from the core, replacing the boundary layer flow entering the receiver holes, which gives rise to the region of high heat transfer.

7 Conclusions

Flow and heat transfer measurements from an experimental study of a preswirl rotor–stator system have been compared with the results of computations. The experiments were conducted on

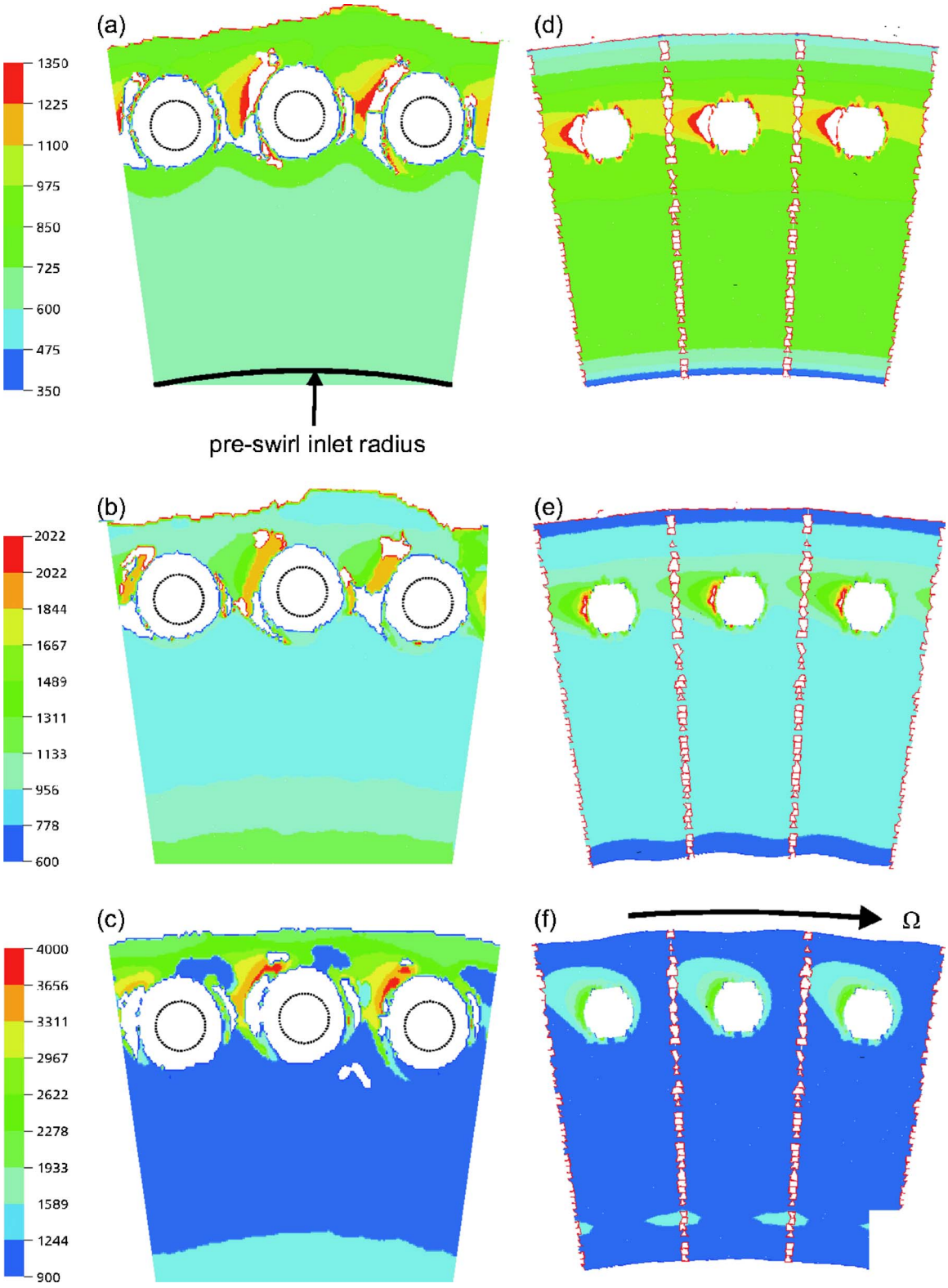


Fig. 7 Experimental (left) and computational (right) Nusselt number contours, $Re_\phi = 0.8 \times 10^6$: (a) $\beta_p = 0.5$, $\lambda_T = 0.13$; (*b*) $\beta_p = 1.0$, $\lambda_T = 0.24$; and (*c*) $\beta_p = 1.5$, $\lambda_T = 0.37$

the "Bath rig:" a purpose-built direct-transfer rig. The measurements were made for $0.8 \times 10^6 < Re_\phi < 1.2 \times 10^6$, $0.1 < \lambda_T < 0.4$ and $0.5 < \beta_p < 1.5$, and the flow structure was considered to be representative of that found in gas turbines. The steady-state computations used a 3D commercial CFD code.

The conclusions based on this rig geometry are as follows:

1. The computed static pressure distribution agrees well with measured values but the tangential velocity, and hence the total pressure, is overpredicted.

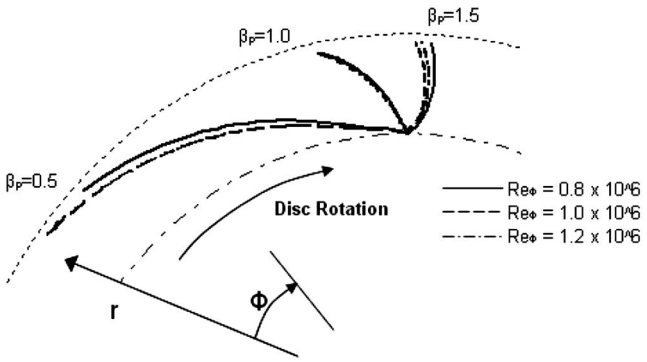


Fig. 8 Streamline plots for flow in the direct route between inlet and outlet

2. The computed values of $\Theta_{b,\mathrm{ad}}$ were in good agreement with a theoretical model for direct-transfer systems.
3. The discharge coefficient, C_D, reached a maximum value when β_1, the core swirl ratio adjacent to the receiver holes, was unity. The computed maximum was $C_D \approx 0.65$, and the experimental maximum was $C_D \approx 0.70$.
4. The computed and measured radial distributions of Nusselt number, Nu, on the rotating disk show evidence of the viscous and inertial regimes. Although Nu tends to increase as Re_ϕ increases, the parameter $\mathrm{Nu}\,\mathrm{Re}_\phi^{-0.8}$ is only weakly dependent on Re_ϕ in the viscous regime. The computations are qualitatively similar to the measurements but, apart from the region near the receiver holes, they do not show good quantitative agreement.
5. The computed and measured contours of Nu show that there is a small region of high heat transfer close to the receiver holes. This is due to the two routes by which flow enters the holes: a "direct" route from the preswirl nozzles and an "indirect" route from the core. The regions of high heat transfer are of importance for designers as they may result in thermal stresses around the receiver holes in turbine disks.

Acknowledgment

Paul Lewis is a Ph.D. student funded by the UK Engineering and Physical Sciences Research Council (EPSRC) through a Doctoral Training Account (DTA). The experimental work was performed by Dr Y. Yan, who was funded by Alstom Power Ltd (now Siemens Industrial Turbines) and EPSRC, and additional discharge coefficient measurements were made by Vinod Kakade from the University of Bath. The authors would like to thank the reviewers for their constructive feedback during the review process.

Nomenclature

a = rotor inner radius
b = rotor outer radius
c_p = specific heat capacity at constant pressure
c_w = nondimensional mass flow rate ($=\dot{m}/\mu b$)
C_D = discharge coefficient for receiver holes
d = preswirl nozzle diameter
G = gap ratio ($=s/b$)
h = heat transfer coefficient
k = thermal conductivity of air
$\dot{m}$ = mass flow rate
M = disk moment
N = number of preswirl nozzles
Nu = Nusselt number ($=q_w r/k(T_w - T_{w,\mathrm{ad}})$)
Pr = Prandtl number ($=\mu c_p/k$)
q_w = rotor wall heat flux
R = recovery factor ($=\mathrm{Pr}^{1/3}$)
Re_ϕ = rotational Reynolds number ($=\rho\Omega b^2/\mu$)
r = radius
r_p, r_b = radii of preswirl nozzles and receiver holes
s = rotor–stator separation distance
T = static temperature
v = velocity
u_τ = friction velocity
x = nondimensional radius ($=r/b$)
y = distance normal to the wall
y^+ = nondimensional wall distance ($=\rho y u_\tau/\mu$)
β = swirl ratio ($=v_\phi/\Omega r$)
γ = ratio of specific heats

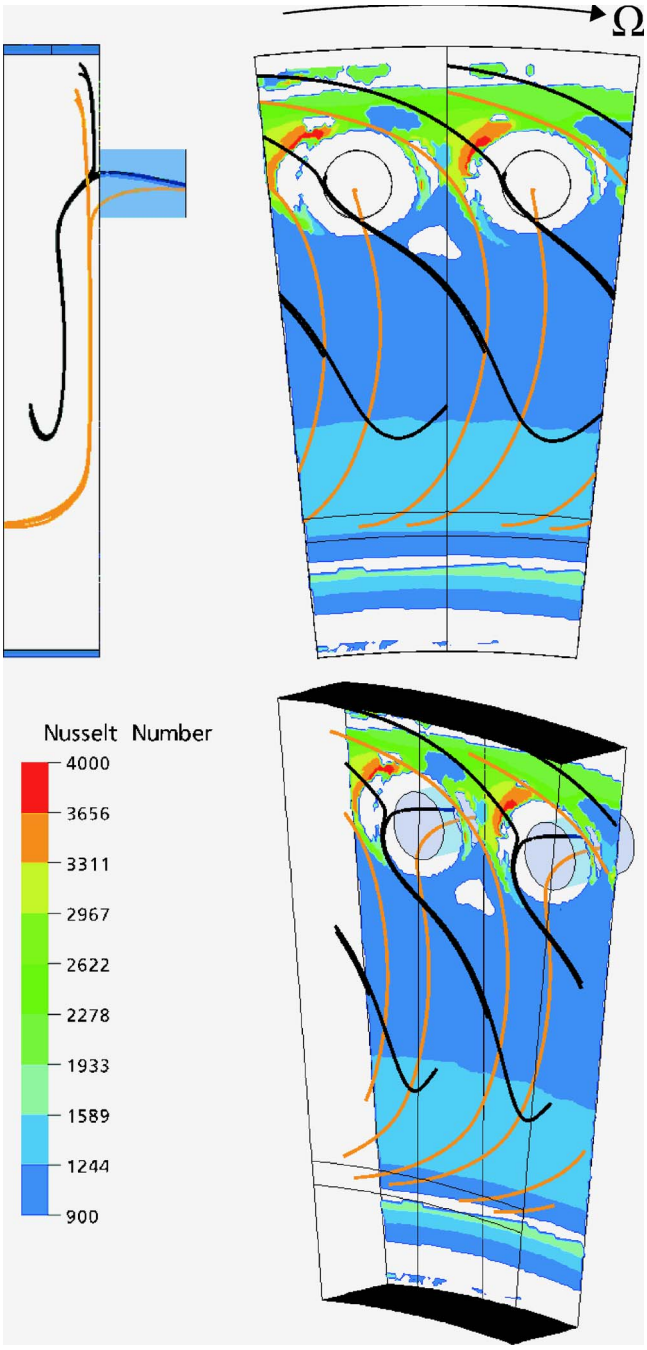


Fig. 9 Computed streamlines superimposed onto experimental heat transfer results: $\mathrm{Re}_\phi = 0.8 \times 10^6$, $\beta_p = 1.5$, $\lambda_T = 0.38$

$\Theta_{b,\mathrm{ad}}$ = adiabatic effectiveness
λ_T = turbulent flow parameter ($=c_w \mathrm{Re}_\phi^{-0.8}$)
μ = dynamic viscosity
ρ = density
Ω = angular velocity of rotor

Subscripts

ad = adiabatic
b = blade-cooling
i = isentropic value
o = total value in stationary frame
p = preswirl
s = stator
t = total value in rotating frame
w = rotor
ϕ, r, z = circumferential, radial, axial direction
∞ = value in core at $z/s=0.5$
1,2 = upstream, downstream locations in a stream tube

References

[1] Meierhofer, B., and Franklin, C. J., 1981, "An Investigation of a Preswirled Cooling Airflow to a Turbine Disc by Measuring the Air Temperature in the Rotating Channels," ASME Paper 81-GT-132.
[2] El-Oun, Z. B., and Owen, J. M., 1989, "Preswirl Blade-Cooling Effectiveness in an Adiabatic Rotor-Stator System," ASME J. Turbomach., **111**, pp. 522–529.
[3] Geis, T., Dittmann, M., and Dullenkopf, K., 2003, "Cooling Air Temperature Reduction in a Direct Transfer Preswirl System," ASME Paper GT2003-38231.
[4] Chew, J. W., Ciampoli, F., Hills, N. J., and Scanlon, T., 2005, "Pre-Swirled Cooling Air Delivery System Performance," ASME Paper GT2005-68323.
[5] Farzaneh-Gord, M., Wilson, M., and Owen, J. M., 2005, "Numerical and Theoretical Study of Flow and Heat Transfer in a Pre-Swirl Rotor-Stator System," ASME Paper GT2005-68135.
[6] Karabay, H., Wilson, M., and Owen, J. M., 2001, "Predictions of Effect of Swirl on Flow and Heat Transfer in a Rotating Cavity," Int. J. Heat Fluid Flow, **22**, pp. 143–155.
[7] Popp, O., Zimmermann, H., and Kutz, J., 1998, "CFD Analysis of Coverplate Receiver Flow," ASME J. Turbomach., **120**, pp. 43–49.
[8] Dittmann, M., Geis, T., Schramm, V., Kim, S., and Wittig, S., 2002, "Discharge Coefficients of a Preswirl System in Secondary Air Systems," ASME J. Turbomach., **124**, pp. 119–124.
[9] Yan, Y., Farzaneh-Gord, M., Lock, G., Wilson, M., and Owen, J. M., 2003, "Fluid Dynamics of a Pre-Swirl Rotor-Stator System," ASME J. Turbomach., **125**, pp. 641–647.
[10] Lock, G. D., Wilson, M., and Owen, J. M., 2005, "Influence of Fluid Dynamics on Heat Transfer in a Pre-Swirl Rotating Disc System," ASME J. Eng. Gas Turbines Power, **127**, pp. 791–797.
[11] Wilson, M., Pilbrow, R., and Owen, J. M., 1997, "Flow and Heat Transfer in a Pre-Swirl Rotor-Stator System," ASME J. Turbomach., **119**, pp. 364–373.
[12] Newton, P. J., Yan, Y., Stevens, N. E., Evatt, S. T., Lock, G. D., and Owen, J. M., 2003, "Transient Heat Transfer Measurements Using Thermochromic Liquid Crystal. Part 1: An Improved Technique," Int. J. Heat Fluid Flow, **24**, pp. 14–22.
[13] Lock, G. D., Yan, Y., Newton, P. J., Wilson, M., and Owen, J. M., 2005, "Heat Transfer Measurements Using Liquid Crystals in a Preswirl Rotating-Disk System," ASME J. Eng. Gas Turbines Power, **127**, pp. 375–382.
[14] Owen, J. M., and Rogers, R. H., 1995, *Flow and Heat Transfer in Rotating-Disc Systems* (Vol. 2, Rotating Cavities), Research Studies Press, Taunton, UK.
[15] Barth, T. J., and Jesperson, D. C., 1989, "The Design and Application of Upwind Schemes on Unstructured Meshes," *Proc. 27th AIAA Aerospace Sciences Meeting*, Reno, NV, January 9–12.
[16] Menter, F. R., 1994, "Two-Equation Eddy Viscosity Turbulence Models for Engineering Applications," AIAA J., **32**(8), pp. 269–289.
[17] Wilcox, D. C., 1998, *Turbulence Modelling for CFD*, 2nd ed., DCW Industries, La Canada, CA.
[18] Ansys Inc., "CFX User Documentation," Version 5.7, Ansys, Inc., Canonsburg, PA.
[19] Kader, B. A., 1981, "Temperature and Concentration Profiles in Fully Turbulent Boundary Layers," Int. J. Heat Mass Transfer, **24**(9), pp. 1541–1544.
[20] Dittmann, M., Dullenkopf, K., and Wittig, S., 2004, "Discharge Coefficients of Rotating Short Orifices With Radiused and Chamfered Inlets," ASME J. Eng. Gas Turbines Power, **126**, pp. 803–808.
[21] Owen, J. M., and Rogers, R. H., 1989, "Flow and Heat Transfer in Rotating-Disc Systems," *Rotor-Stator Systems*, Research Studies Press, Taunton, UK, Vol. 10.
[22] Luo, X., Zhang, C., Xu, G., Tao, Z., and Ding, S., 2004, "Measurements of Surface Temperature Distribution on a Rotating Disk With Blade Cooling Holes Using Thermochromic Liquid Crystal," ASME Paper GT2004-53519.

Water Injection Effects on Compressor Stage Operation

I. Roumeliotis
Research Assistant
e-mail: jroume@ltt.ntua.gr

K. Mathioudakis
Associate Professor

Laboratory of Thermal Turbomachines,
National Technical University of Athens,
Iroon Polytechniou 9,
Athens 15773, Greece

In the present paper, experimental work concerning the effect of water injection on a compressor stage is presented. The effect on compressor stage performance and stability is examined for water injection up to 2%. The behavior of the airflow in the blade rows is examined through aerodynamic measurements. The results indicate that although the water injection appears to not have any significant effect on the flow pattern and to stage pressure rise and stall margin, there is a measurable effect on compression efficiency, which seems to result mainly from losses of a mechanical nature and water acceleration. The efficiency degradation is proportional to the water ratio entering the engine.
[DOI: 10.1115/1.2718223]

Keywords: water injection, wet compression, two-phase flow

Introduction

The increasing use of gas turbines in the power generation industry has created an additional incentive for the further improvement of their performances. In the last few years, several techniques have been proposed for gas turbine power and efficiency augmentation, such as steam or water injection into the combustion chamber or the airflow, with much interest in the moisture air cycle [1,2] and evaporative gas turbine [3]. The fact that gas turbine output and efficiency drop during high ambient temperature periods, when demand usually increases, has led to the broad application of inlet air cooling [4]. Overspray of water droplets inside the compressor has found an increasing application.

For the effect of water injection to the overall engine performance, experimental results have been presented by Utamura et al. [5], Cataldi et al. [6], and Deneve et al. [7].

Although a significant attention has been drawn on inlet injection, little experimental work has been undertaken in order to understand the effect of water injection to the stage operation, rather than the effect on a whole engine. In the case of a multistage compressor, the interstage evaporation is the dominant factor resulting in stage rematching (see, for example, [8,9]). In order to further understand the effect of water injection on compression, the effect on a single stage should be isolated.

The single-stage operation analysis allows further understanding of the effect of water injection on the compressor process, mainly to examine the effect of droplet-laden flow on stage behavior as well as to give some guidelines for the modeling of the process. The main effect of concern is the influence of water droplet presence inside the stage on the performance characteristics of the stage, namely, the pressure-rise coefficient and efficiency.

In the present work, the analysis of the behavior of a single compressor stage has been undertaken. The stage is rotating at low speed, allowing the assumption of incompressible flow (Mach number is below 0.3 throughout the stage). Because the temperature rise is small, the effect of droplet evaporation during compression may be neglected as a first approximation, and thus solely the aerodynamic and mechanical effects of water injection on the compression process can be examined.

This configuration allows detailed aerodynamic measurement of the effect of water injection on stage performance parameters, and the compressor flow field.

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received June 12, 2006; final manuscript received August 28, 2006. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2006; Land, Sea and Air (GT2006), May 8–11, 2006, Barcelona, Spain. Paper No. GT2006-90427.

The measurement techniques adopted in order to have reliable measurements for the stage performance parameters are discussed. As pointed out by other researchers [10,11], special care for measurements in droplet-laden flows has to be taken.

Experimental Setup

The setup of the experimental investigation and the procedure followed was dictated by the main purpose of the investigation, which was to examine the effect of water injection on compressor operation characteristics, namely, performance parameters and stability. The facility used and the choice of instrumentation are presented in the following.

Test Facility. For the experiments, a one-stage axial compressor was used. The stage is designed to produce a pressure ratio of 1.57 with a mass flow rate of 27.5 kg/s for a rotational speed of 18,000 rpm. Tests conducted were at low rotating speed, up to 50% of the nominal speed.

The compressor is installed, and the experiments were carried out on the high-speed test rig of the Laboratory of Thermal Turbomachines, National Technical University of Athens, modified to operate with wet conditions.

Because of restrictions applied by the inlet geometry and the test rig configuration, a manifold of a maximum of 44 impaction pin nozzles of Mee Industries were used. The water flow is determined by the number of nozzles and the operating pressure. Thus by varying the number of nozzles and the supply pressure, a range of water-injected ratio can be obtained. The usable number of nozzles dictated the maximum water-to-air ratio to be at 2% for low-speed operation. This is an acceptable water ratio because in most applications, overspary seldom exceeds 1% of the airflow.

The nozzles are axisymmetricaly distributed at the bellmouth inlet. In Fig. 1, the fog manifold and the stage are shown. Because of the nature of the experiment, special care should be given to the test-facility configuration. The initial inlet duct was a short bellmouth. In order to examine whether the inlet configuration was suitable for water injection, a droplet-trajectories analysis was undertaken. The droplet trajectories analysis was done solving the Lagrangian equation of motion for drag force and gravity force in correlation with a meridian flow solver [12] and incorporating the droplet evaporation model proposed by Chaker et al. [13].

The results indicated that for the original configuration and for the specific injection position, the bigger droplets will impinge on the bellmouth walls. This is due to the sudden change of bellmouth diameter and to the small length of the initial design; thus, the droplets do not have the time to acquire a velocity close to the flow velocity. If a large number of droplets impinge on the walls,

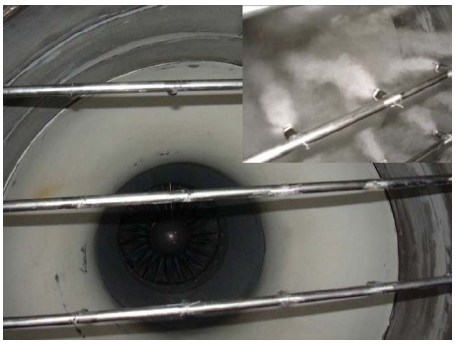


Fig. 1 Compressor stage and nozzle manifold

then a sheet of water will be created and ingested by the compressor, thus creating an abnormal effect. For this reason, the initial bellmouth was elongated by 0.6 m.

An indicative result of the trajectories analysis is presented in Fig. 2 for the bellmouth initial configuration and the elongated configuration, assuming a droplet initial velocity of 120 m/s, droplet diameter of 30 μm, and air mass flow of 13.6 kg/s. It is evident that placing the nozzles further downstream reduces the risk of creation of liquid film prior to rotor.

Measurements for Performance Evaluation. The compressor test rig is fitted with instrumentation allowing measurement of the quantities necessary to derive the compressor performance map. A schematic diagram of the test setup is given in Fig. 3. The nozzle manifold is positioned 1.6 m from the stage inlet.

During testing the following measurements were acquired:

1. Flow rate of injected water
2. Pressure delivered at nozzle manifold
3. Compressor rotating speed
4. Power of the shaft
5. Ambient temperature, pressure, and relative humidity

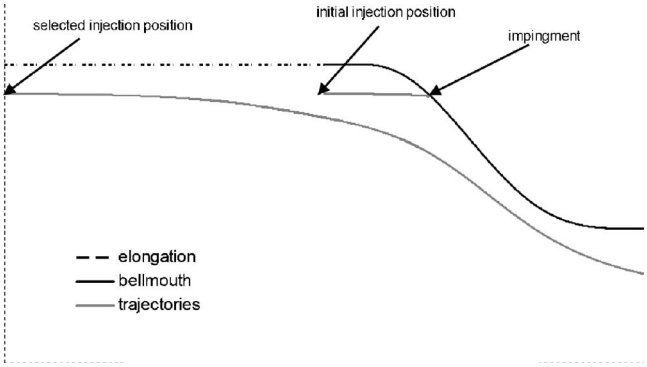


Fig. 2 Droplet trajectories calculation for inlet bellmouth

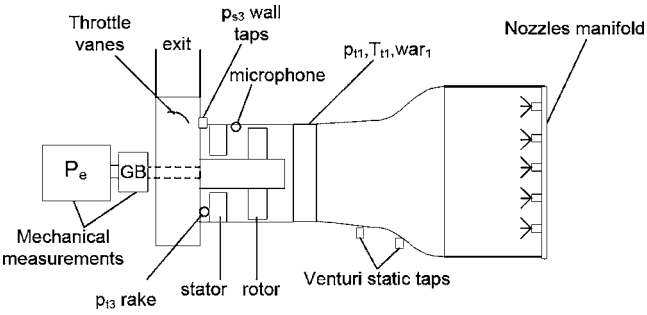


Fig. 3 Experimental setup

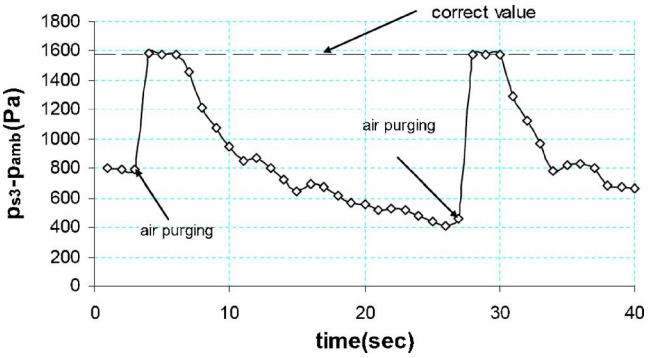


Fig. 4 Effect of measurement tubing plugging on stage exit static pressure and effect of air purging

6. Compressor inlet total pressure by using a Kiel-type probe with a 3 mm hole
7. Compressor inlet static pressure using eight wall taps
8. Inlet flow rate is measured using the bellmouth as Venturi flowmeter. The static wall taps used (eight at each position) where positioned close to the neck of the bellmouth where the evaporation is minimal
9. Compressor inlet absolute humidity and temperature measured by bleeding air from the compressor inlet
10. Static pressure at the stage exit by using eight static pressure tappings of 1.5 mm diam
11. Total pressure at the stage exit by using a total pressure rake consisting of five sensors
12. Gearbox and compressor bearings inlet and outlet oil temperature along with the oil flow

The power consumed by the compressor is calculated by using the power on the shaft, subtracting the parasitic losses. Because of water droplet presence, the measuring quantities were taken with specific techniques in order to avoid water contamination.

The static pressure wall taps were 1.5 mm diam. In order to be sure that no blockage due to water droplets or water sheet may occur during the measurements, all the pressure wall taps were equipped with a purge system. The static pressure tapings were continuously flushed with air, except for the time of measurement as recommended by Day et al. [11].

During testing, if no flushing was used, plugging was evident at the pressure tapping, even when no water was visible inside the pressure tubes. In Fig. 4, the measuring compressor exit static pressure is presented without continuous cleaning. The drop of static pressure is evident after 3 s of measuring. When the tapings were flushed, the measurement is returning to the initial value.

In order to obtain aerodynamic data from the wet compressor, a three-hole probe was designed with enlarged holes of 1.5 mm. The three-hole probe consisted of three individual tubes of 2.5 mm diam assembled together. The left and right holes are at 90 deg angle and both are at a 45 deg angle from the center hole. The three-hole probe was calibrated through a known flow field and displayed very good agreement with the expected behaviour. A Kiel-type total pressure probe with a 3 mm hole was also designed and manufactured. The cleaning system was applied to the three-hole probe and to the Kiel-type total pressure probe.

When small droplets are injected in front of the compressor, evaporation occurs into the bellmouth, changing the inlet temperature and air synthesis and, consequently, inlet density. In order to acquire reliable results concerning the performance parameters, the stage inlet temperature and absolute humidity must be measured as accurately as possible.

The stage inlet humidity is measured by bleeding air from the compressor inlet to a chamber equipped with a relative humidity sensor and a Pt100 temperature sensor. The bleeding is done by

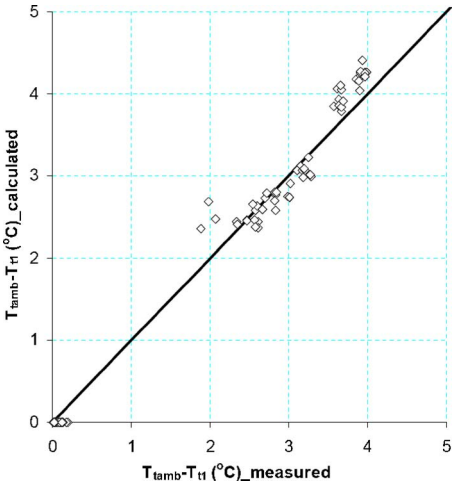


Fig. 5 Comparison of measured versus calculated via water to air ratio stage inlet temperature

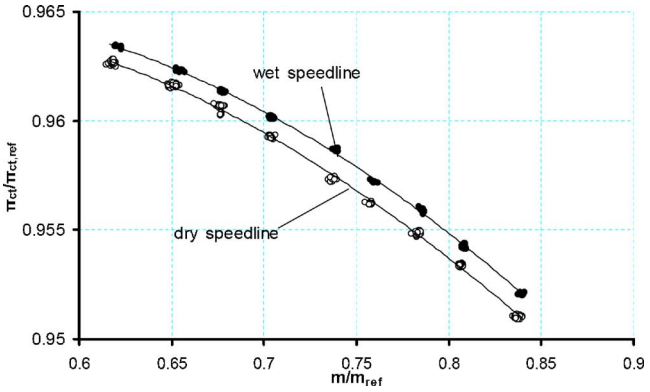


Fig. 6 Stage map for dry and wet operation (1% water ratio)

using a bent tube facing downstream. The water-to-air ratio at the compressor inlet is calculated using relative humidity, temperature, and pressure measurements at the humidity chamber.

The compressor inlet temperature was measured using the same technique by a Pt 100 Class A temperature sensor. The temperature sensor is positioned within the tube feeding the humidity chamber, immediately after the connection point to the casing.

The nonintrusive method was selected, although the use of a shielded temperature sensor was examined. The shielded sensor was prone to water contamination, thus indicating the wet bulb temperature.

In Fig. 5, the stage inlet temperature as measured by the temperature sensor and as calculated using the ambient and measured stage inlet water-to-air ratio is presented. The temperature decrease is in good agreement using both measurements.

Having a reliable measure for the inlet conditions allows the calculation of inlet density, which is a critical factor for the calculation of stage performance parameters, as well as to have correct results from the Venturi-type inlet mass measurement [14].

The inlet density is calculated using the measured quantities $(p_{t1}, T_{t1}, p_{s1}, w_{ar_1})$, in addition to the calculated static temperature. The static temperature is calculated by applying isentropic process, only for the gas component of the flow.

Measurements for Unstable Operation Determination. In order to have a reliable indicator of unstable operation, fast response measurements should be used. Such an indicator is the measurement of unsteady flow velocity using hot wires. Thus in the case of dry measurements, one hot wire was placed at the rotor exit. The hot wires used in this study were of the type 55P11 with Pt-plated tungsten wire, manufactured by Dantec.

The hot wires are very sensitive sensors; thus, the use of them in droplet-laden flows is prohibitive. As presented by Aretakis et al. [15] stall can be detected by acoustic measurements. In order to have an unstable operation criterion for wet operation, the sound emitted by the compressor was measured.

Two microphones were used, one positioned outside of the compressor facing the casing at the rotor exit and one positioned at the compressor casing through a hole at the rotor exit. The microphones were the condenser type, with a frequency range of 0–14 kHz, battery operated. The fast response measurements were derived from data strings of 8192 samples per channel, with a sampling frequency of 30,000 Hz.

Effect of Water Injection on Compressor Performance

For the following results, the operating pressure of the nozzle manifold is 140 bar, resulting in a D_{v05} diameter of ~ 12 μm [4], except if stated otherwise. The droplet diameter is given as an indicator only because no droplet diameter measurements were undertaken, and as pointed out by Chaker [16], there are several parameters that influenced the diameter of the droplets entering the engine.

A water-injection well-known initial effect is to lower stage inlet temperature thus leading the stage to higher corrected speed. The dry map of the stage, along with the map for 1% of water injection, is presented in Fig. 6, using uncorrected magnitudes, for the case of constant mechanical speed.

The expected change due to inlet evaporation on pressure ratio and flow mass rate is observed; as for the case of water injection, the stage is working on higher corrected speed. The mass flow rate is increasing, although little for the full open throttling position, and the pressure ratio is also increasing.

Effect on Stage Characteristics. In order to fully evaluate the effect of water injection on the stage performance characteristics, when no evaporation occurs inside the stage, the nondimensional flow and pressure-rise coefficients should be used. All calculations are done according to the stage-measured inlet conditions $(T_{t1}, w_{ar_1}, p_{t1}, p_{s1})$; thus, the used pressure rise coefficient is

$$\Psi_{t-t} = \frac{p_{t3} - p_{t1}}{\rho_1 \cdot U_{\mathrm{TIP}}^2} \tag{1}$$

The same approach is used for the total to static pressure rise coefficient. The effect of nonevaporated water on the air density is not taken into consideration due to the negligible volume occupied.

The effect of water injection up to 2% of the airflow appears to have no measurable effect on the total-total and total-static pressure-rise coefficient. This observation comes in agreement with the observations of Day et al. [11] for low water quantities and with the fact that the nondimensional parameters do not change significantly due to humidity. The total-static and total-total pressure-rise coefficient for a fixed injected water quantity, resulting in a water ratio varying from 1.1% at full open throttling position to 1.5% at the maximum pressure-rise coefficient point is presented in Figs. 7 and 8, respectively.

Effect on Stall Margin. In Fig. 9, the dry and wet characteristic for a maximum ratio of 1.5% is presented. The effect of quantities of water up to 2% ratio seems to have no measurable effect on the operating point where initial stall occurs. The microphones presented good behavior for the case of unstable operation.

Even if the stage stall limits do not change, the whole compressor stall limit and stall margin will be negatively influenced by water injection, as pointed out by Roumeliotis and Mathioudakis [17]. This reduction on the surge line is due to the constant evaporation inside the compressor, resulting in a significant rematching due to density change—an effect not present in the experimental

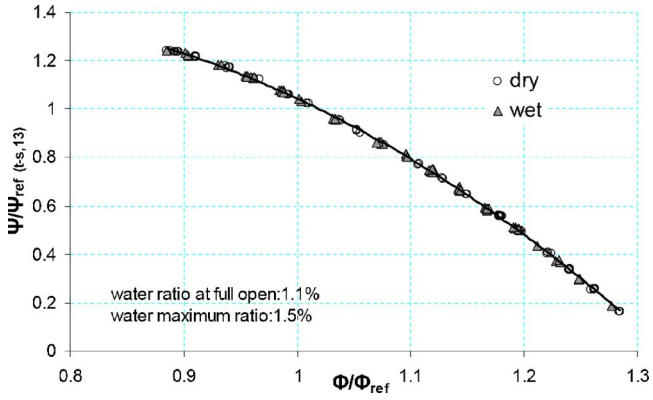


Fig. 7 Total static pressure rise coefficient for water injection up to 1.5%

configuration under investigation. The droplet diameter as controlled by the operational pressure of the manifold seems to have no additional effect on the pressure-rise coefficient and stage stall limit for manifold operational pressure down to 50 bar (D_{v05} $=20$ μm [4]).

Detailed Aerodynamic Measurements. As presented above, injection of a modest amount of water appears to have no effect on the pressure-rise coefficient and on the stage surge margin. In order to further examine the validity of this observation, more detailed measurements of the aerodynamic behavior of the compressor were undertaken.

First, the inlet pressure profile was measured. The measurements were done using the Kiel-type total pressure probe, equipped with a cleaning system, allowing water-free measurements for the whole inlet annulus. In Fig. 10, the inlet total pressure variation for 1.5% of water injection is presented. As one can see, the inlet total pressure profile does not change due to water injection.

The geometry of the tested compressor is not allowing the full traversing of the rotor blade span. The close proximity between rotational and stationary blades did not allow further insertion of the probe. The 60% of the blade span can be measured herein. As reported by Day et al. [11], a significant deviation of the rotor exit magnitudes occurs for nontrivia water-injected quantities, which is evident through the whole span; thus, the measurements that can be acquired herein can give sufficient information if any deviation due to water injection occurs.

In Fig. 11, the rotor exit normalized relative angle turn is presented for the case of 2% of injected water ratio. No significant deviation is observed for the rotor relative exit angle. Thus, no

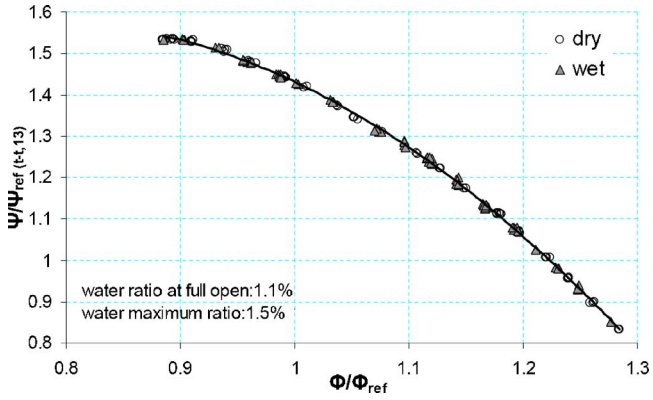


Fig. 8 Total-total pressure rise coefficient for water injection up to 1.5%

significant change is expected on the blade operation. Further justification for this conclusion can be acquired using the total pressure loss coefficient. The total pressure loss coefficient is defined as

$$\bar{\omega} = \frac{p_{tR1} - p_{tR2}}{(1/2)\rho_1 W_1^2} \tag{2}$$

$$p_{tRi} = p_{si} + \rho_1 \frac{W_i^2}{2} \tag{3}$$

In Fig. 12, the total pressure loss coefficient is presented for the case of 2% water injection.

The pattern of absolute exit flow angle from the stator is also unaffected by the water injection, as one can see in Fig. 13, where the stator absolute angle is presented for 1.3% injected water ratio. Summarizing the effect of the injection of a modest amount of water inside the compressor appears to have no significant effect on pressure rise and stall limit.

Detailed aerodynamic measurements have shown that no measurable effect on blade operation occurs due to water injection. Although the measurements close to casing must be treated with caution.

Effect on Compressor Efficiency. Water injection results in a power increase. In Fig. 14, the power increase due to water injection for a full characteristic and two different steady quantities of injected water mass flow is presented. The power presented is corrected for inlet conditions and rotating speed, as proposed in AGARD [14] for inlet humidity. The consumed power by the compressor is increasing as the injected water ratio increases as expected. In Fig. 15, the direct measured power increase is presented for a throttling position. The increase of consumed power seems to be linear with the quantity of water entering the stage. The power increase due to water injection is overestimated because water injection results in increased mass and pressure ratio for same throttling positions.

As the pressure ratio increases with water injection due to inlet evaporation, the behavior of the consumed power does not give a clear image of the effect of water injection on compression efficiency. Thus, the compressor efficiency is selected as the magnitude to be examined. The compressor efficiency with the assumption of incompressible flow inside the compressor (a valid assumption because the Mach number is below 0.3) can be obtained directly by measurements, using the following relation:

$$\eta_{C,is} = \frac{\Delta p_{t-t,13}/\rho_1}{P_{\text{comp}}/\dot{m}_{\text{in}}} \tag{4}$$

In Eq. (4), the magnitudes $\Delta p_{t-t,13}$, P_{comp}, and $\dot{m}_{\text{in}}$ are directly measured quantities and ρ_1 is calculated according to the measured inlet conditions.

The stage is operating at low speed, where the temperature rise is relatively small (maximum temperature rise for the operating speeds of concern is 12 deg); thus, any additional evaporation inside the stage, as the air is entering with increased humidity, can be neglected as a first approach. Initial calculations with a one-dimensional model [17] indicate that the amount of heat exchange between water and air is of very small magnitude for the test conditions and can thus be neglected in the present calculations.

In Fig. 16, the percentage of efficiency decrease for various water injection quantities and operation points is presented. The decrease has been calculated by using the polynomial approach of the dry isentropic efficiency (Φ-η_{cis}). The raw values for the wet compression have been used (no polynomial approach).

The injected quantity percentage was controlled by changing the rotating speed and the operating pressure of the nozzle manifold. No significant effect from the variation of the manifold operating pressure (50–140 bar) was observed; thus, no significant

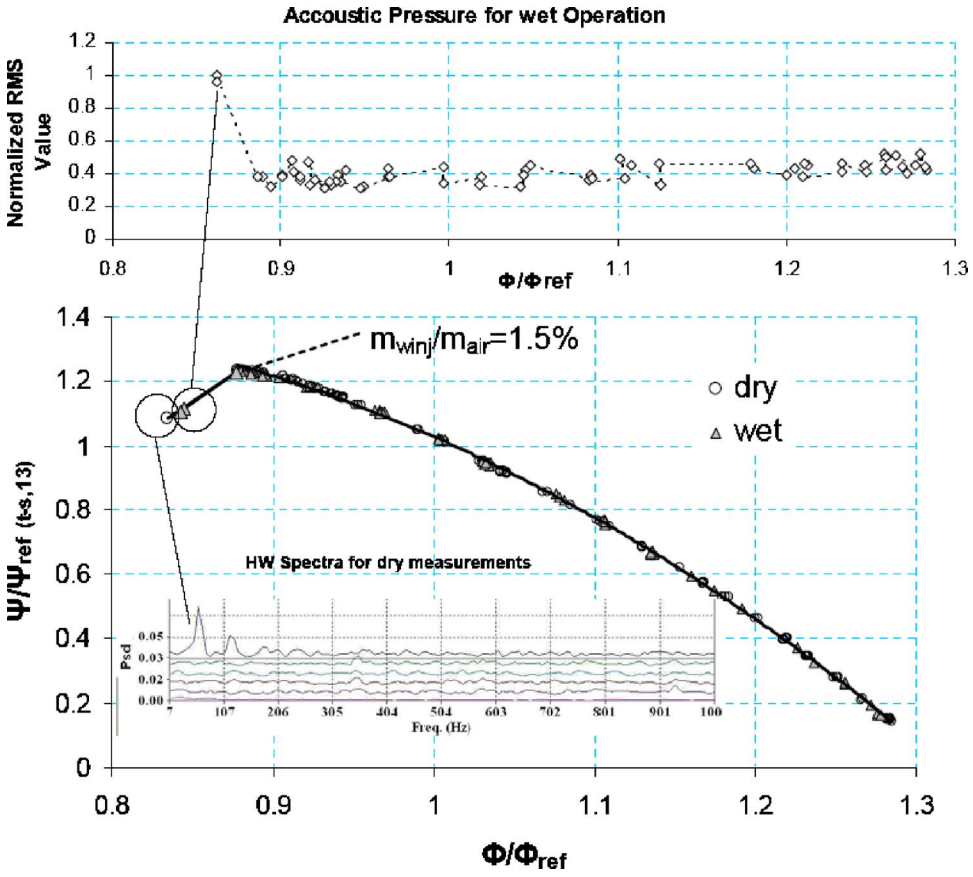


Fig. 9 Effect of water injection on stall margin for 1.5% of injected water ratio

effect from the variation of the corresponding droplet diameter is expected. The trend of the efficiency decrease due to water injection seems to be linear with the water quantity entering the compressor.

Arsen'ev and Berkovich [18] have proposed that the mechanical losses due to moisture (droplet acceleration, droplet breaking) can be taken into consideration for first-stage injection by incorporating 1% losses for every percentage of water injected.

The experimental results for the efficiency decrease suggest that the losses are at that range (line slope: 1.1) and proportional to the water quantity entering the stage.

As pointed out in the previous section, no significant changes in the flow is evident; thus, the additional losses resulting in an efficiency decrease should be searched on the water acceleration and rotor braking due to droplet impact. Further investigation of the reasons for the efficiency decrease should be undertaken in order to clearer quantify the effect.

Conclusions

The effect of water injection on a compressor stage has been isolated in order to investigate the effect due to droplet interaction with the stage. The tests were conducted at low rotational speed in order to avoid any evaporation inside the stage. The evaporation effect at the inlet has been taken into consideration. The special techniques used in order to acquire measurements with wet compression perform well; thus, reliable measurements can be obtained from a wet stage.

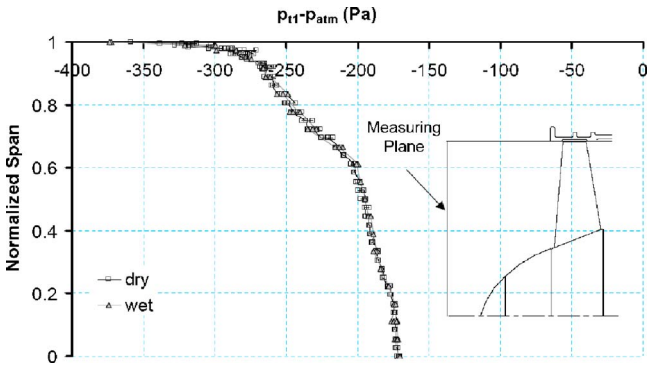


Fig. 10 Compressor inlet pressure profile for dry and wet operation (water ratio 1.5%)

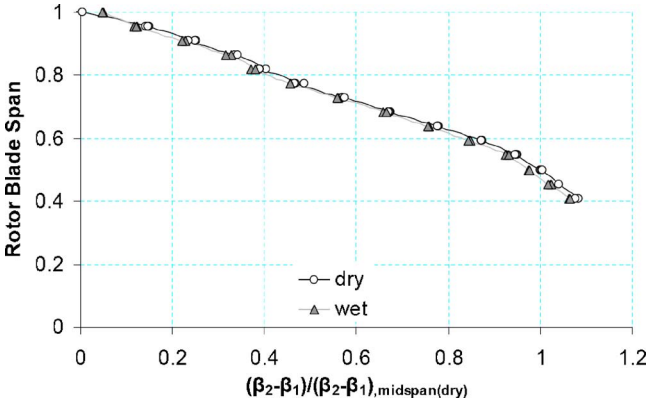


Fig. 11 Rotor relative angle turn for 2% water injection ratio

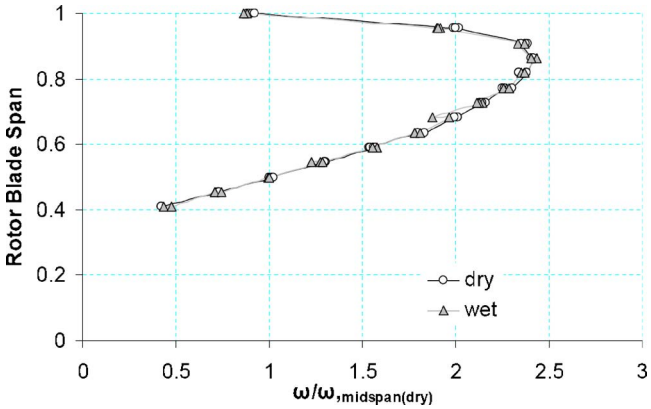


Fig. 12 Total pressure loss coefficient for 2% of injected water ratio

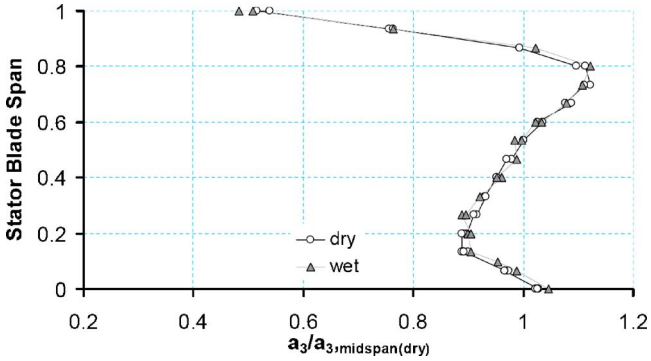


Fig. 13 Stator absolute exit angle for 1.3% of injected water

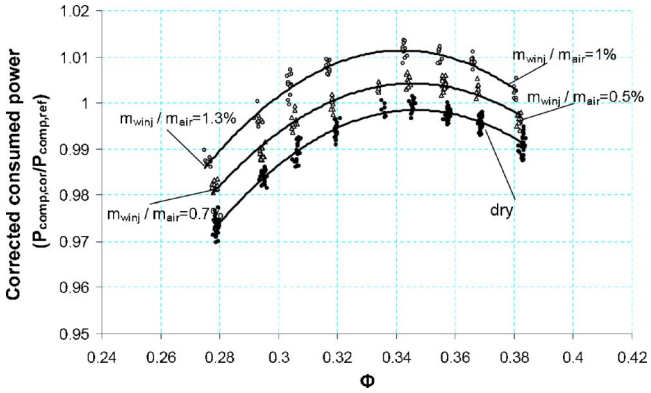


Fig. 14 Corrected consumed power for dry and wet operation

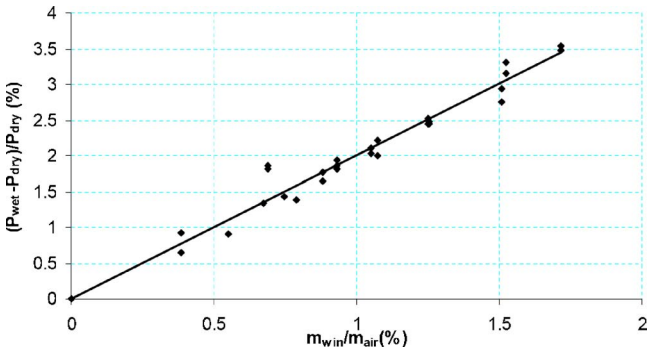


Fig. 15 Increase of consumed power versus inlet water ratio for a throttling position

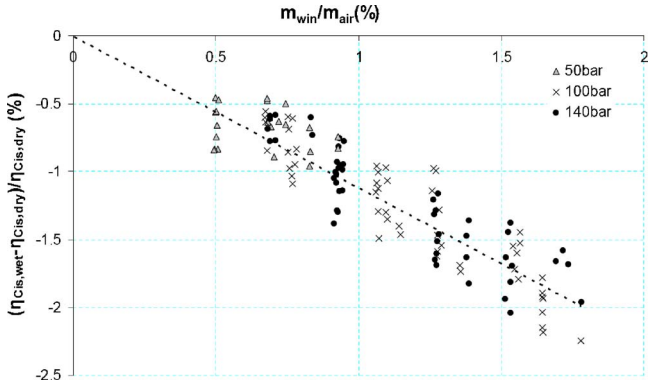


Fig. 16 Decrease of compression isentropic efficiency versus stage inlet water ratio

For a water injection quantity up to 2% of the airflow, no measurable effect on the pressure-rise coefficient was observed. The compressor stability was unaffected by water injection because the operating point where the stall occurs presents no significant change.

Detailed aerodynamic measurements indicate that no significant effect on blade operation occurs due to water injection because no redistribution of the flow was measured.

The effect of water injection on consumed power is evident and has been shown to increase linearly with the amount of water entering the stage. The effect of water injection on efficiency is also evident because a decrease on compression efficiency was measured. The decrease presented linearity with the water quantity entering the compressor.

Further understanding of the mechanism affecting the efficiency drop due to water injection is needed. The results indicate that because the flow is not significantly altered, water acceleration and rotor braking due to droplet impingement may be the reason.

Acknowledgment

The work for this paper has been carried out in the frame of the HERAKLITOS program. The program is cofunded by the European Social Fund (75%) and National Resources (25%). The cooperation of Mee Industries, and T. Mee personally, in purchasing the manifold is also gratefully acknowledged.

Nomenclature

D_{v05} = mass median droplet diameter
m = mass flow rate
p = pressure
P_{comp} = power consumption by the compressor
P_e = electrical power
T = temperature
U = blade speed
V_a = axial velocity
W = relative velocity
w_{ar} = water to air ratio
α = absolute angle
β = relative angle
η_{Cis} = stage isentropic efficiency
π_C = compressor pressure ratio
ρ = density
$\Phi = V_a/U_{\text{tip}}$ = mass flow coefficient
Ψ = pressure rise coefficient
ω = total pressure loss coefficient

Subscripts

amb = ambient conditions
air = gaseous mixture at stage inlet
R = relative magnitude

ref = reference value
s = static magnitude
t = stagnation magnitude
tip = blade tip
winj = injected water
win = water at stage inlet (e.g., injected quantity minus evaporated at stage inlet duct)
1 = rotor inlet
2 = rotor exit
3 = stator exit

References

[1] Utamura, M., Takehara, I., and Karasawa, H., 1998, "MAT, A Novel, Open Cycle Gas Turbine for Power Augmentation," Energy Convers. Manage., **39**(16–18), pp. 1631–1642.

[2] Utamura, M., Takehara, I., Horii, N., and Kuwahara, T., 1997 "A New Gas Turbine Cycle for Economical Power Boosting," ASME Paper No. 97-AA-142.

[3] Horlock, J. H., 1997, "The Evaporative Gas Turbine [EGT] Cycle," ASME Paper No. 97-GT-408.

[4] Meher Homji, C. B., and Mee, III, T. R., 2000, "Inlet Fogging of Gas Turbine Engines—Part A: Theory, Psychrometrics and Fog Generation," ASME Paper No. 2000-GT-307.

[5] Utamura, M., Kuwahara, T., Murata, H., and Horii, N., 1999, "Effect of Intensive Evaporative Cooling on Performance Characteristics of Land Based Gas Turbine," PWR-Vol. 34, ASME Joint Power Generation Conference, Vol. 2.

[6] Cataldi, G., Güntner, H., Matz, C., McKay, T., Hoffmann, J., Nemet, A., Lecheler, S., and Braun, J., 2004, "Influence of High Fogging Systems on Gas Turbine Engine Operation and Performance," ASME Paper No. GT2004-53788.

[7] Deneve, M., De Tandt, B., Cornelis, N., Bultereys, C., and Gijbels, S., 2005, "Results of the First Application of the Swirlflash Wet Compression System on a 150 MW Heavy Duty Gas Turbine," ASME Paper No. GT2005-68726.

[8] White, A., and Meacock, A., 2003, "An Evaluation of the Effects of Water Injection on Compressor Performance," ASME Paper No. GT 2003-38237.

[9] Bagnoli, M., Bianchi, M., Melino, F., Peretto, A., Spina, P. R., Bhargava, R., and Ingistov, S., 2004, "A Parametric Study of Interstage Injection on GE Frame 7EA Gas Turbine," ASME Paper No. GT2004-53042.

[10] Tsuchiya, T., 1982, "Aerothermodynamics of Axial-Flow Compressors With Water Ingestion," Ph.D. thesis, Purdue University.

[11] Day, I., Freeman, C., and Williams, J., 2005, "Rain Ingestion in Axial Flow Compressor at Part Speed," ASME Paper No. GT-2005-68582.

[12] Kiousis, P., Chaviaropoulos, P., and Papailiou, K. D., 1992, "Meridional Flow Calculation Using Advanced CFD Techniques," ASME Paper No. 92-GT325.

[13] Chaker, M., Meher-Homji, C. B., and Mee, III, T., 2002, "Inlet Fogging of Gas Turbine Engines—Part A: Fog Droplet Thermodynamics, Heat Transfer and Practical Considerations," ASME Paper No. GT-2002-30562.

[14] AGARD-AR-332, 1995, "Recommended Practices for the Assessment of the Effect of Atmospheric Water Ingestion on the Performance and Operability of Gas Turbines Engines," North Atlantic Treaty Organization.

[15] Aretakis, N., Mathioudakis, K., Kefalakis, M., and Papailiou, K., 2004, "Turbocharger Unstable Operation Diagnosis Using Vibroacoustic Measurements," ASME J. Eng. Gas Turbines Power, **126**, pp. 840–847.

[16] Chaker, M., 2005 "Key Parameters for the Performance of Impaction—Pin Nozzles Used in Inlet Fogging of Gas Turbines," ASME Paper No. GT2005-68346.

[17] Roumeliotis., I., and Mathioudakis, K., 2006, "Evaluation of Interstage Water Injection Effect on Compressor and Engine Performance," ASME J. Eng. Gas Turbines Power, **128**, pp. 849–856.

[18] Arsen'ev, L. V., and Berkovich, A. L., 1996, "The Parameters of Gas Turbine Units With Water Injected Into the Compressor," Thermal Engineering, **43**(6), pp. 461–465.

Development, Testing, and Implementation of a Gas Turbine Starting Clutch With Manual Turning Feature for U.S. Navy Ships

Morgan L. Hendry
SSS Clutch Company, Inc.,
610 West Basin Road,
New Castle, DE 19720
e-mail: engineering@sssclutch.com

Matthew G. Hoffman
Surface Combatant Gas Turbine Engineering Branch,
NAVSEA Philadelphia,
Code 934,
Philadelphia, PA 19112
e-mail: matthew.g.hoffman@navy.mil

Most gas turbine generators rely on an automatic-engaging, free-wheel clutch to connect a starting motor to accelerate the gas turbine generator from zero to some intermediate speed to enable ignition and then provide torque assistance to a higher speed until the gas turbine is self-sustaining. The U.S. Navy has used various designs of starter motors and clutches for its gas turbine fleet. In addition, there has been a requirement to periodically borescope each gas turbine, which has necessitated removal of the starting system and clutch assembly in each instance. This paper examines the U.S. Navy experience with starting clutches and provides details of the development and testing of a synchronous-self-shifting clutch with an additional, stationary, manual turning feature to provide very slow and precise gas turbine rotor rotation for borescope purposes. This paper also gives details of the installation of the first two prototype clutches on the USS Ramage, DDG 61, operating experience for approximately four years, and possible future installations of this type of clutch in U.S. Navy gas turbine generator sets.
[DOI: 10.1115/1.2434347]

Background

Since the mid-1970s, the U.S. Navy has used five versions of Ship Service Gas Turbine Generators (SSGTG) onboard each of its destroyers and cruisers to provide for all shipboard electrical power required for machinery, weapons, and hotel services. Over 350 SSGTGs have been produced for use onboard DD 963 Spruance Class Destroyers, DDG 993 Kidd Class Guided Missile Destroyers, CG 47 Ticonderoga Class Guided Missile Cruisers, and DDG 51 Arleigh Burke Class Destroyers. These SSGTGs have been redesigned, upgraded, and uprated through the years to improve maintenance and reliability and to meet the changing and increased power requirements of later ships. This paper will focus primarily on the SSGTGs produced for the DDG 51 Class ships.

The DDG 51 Class ships utilize a Rolls Royce Corporation (RRC), formerly Allison, Model 501-K34 gas turbine, which connects to an AC generator through a 7.6781 ratio, single-stage, vertically offset, double-helical reduction gearbox as shown in Fig. 1. This particular SSGTG, Model AG9140RF, is used by the U.S. Navy to provide electrical power on DDG 79 through DDG 112 ships. All the SSGTGs prior to DDG 79 use a pneumatic starter as the only means of starting the gas turbine, and it is cantilever-mounted using a quick disconnect Marmon clamp on the generator side of the reduction gear.

The starter is an air turbine starter that is connected to the high-speed pinion opposite the gas turbine with a mechanical clutch. A number of different starters and clutches have been used on these various SSGTGs, as shown in Fig. 2. Normal gas turbine starting is a cross-bleed start using bleed air from the discharge of the fourteenth stage of the compressor from a gas turbine of another SSGTG that is already running. Because the SSGTGs must also be capable of a dark ship start (no bleed air available), the U.S. Navy has historically relied on a shipboard, high-pressure air system consisting of 20.7 MPa (3000 psi) air stored in bottles. This air is supplied to the starter through high-pressure piping and a pressure-reducing manifold.

Because of the high cost of maintenance, low reliability, weight reduction, space savings, and concern for safety in manned spaces in the ship, an alternative starting system was designed in the mid-1990s to replace the high-pressure air starting system in these SSGTGs. This was done to ease the duty on the primary starter by having a separate means for a dark ship start.

The Redundant Independent Mechanical Start System (RIMSS) consists of an RRC Model 250-KS4 gas turbine (an aeroderivative of the helicopter 250-C20 gas turbine) that is mounted inside the SSGTG module over the reduction gear and is connected to the high-speed reduction gear pinion (between the air turbine starter and reduction gear pinion) by a 2.762:1 ratio, parallel-shaft, transfer gearbox. This transfer gearbox contains a size 30 T synchronous-self-shifting (SSS) clutch to automatically connect/disconnect the Model 250 engine and transfer gearbox to the reduction gear, as shown in Fig. 3. The first installation was on the USS Porter (DDG 78).

In order to facilitate maintenance of the RRC Model 501 gas turbine, the U.S. Navy periodically removes the air turbine starter to enable the high-speed pinion to be manually rotated at a slow speed so that the Model 501 gas turbine can be borescoped and inspected. The Marmon clamp mounting enables the air turbine starter to be removed and reinstalled by two men in about an hour, but frequent removal and reinstallation have led to premature starter motor and clutch problems, particularly on the DD963 and CG 47 Class ships.

The Navy has had a long experience with variants of the air turbine starter that ultimately resulted in unacceptable start performance, high cost, and reduced mean time between failure (MTBF). Since their early use on the DD 963 Spruance Class destroyers, the air turbine starters have been problematic for a

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received June 14, 2006; final manuscript received July 26, 2006. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2006: Land, Sea and Air, May 8–11, 2006, Barcelona, Spain. Paper Number GT2006-91256.

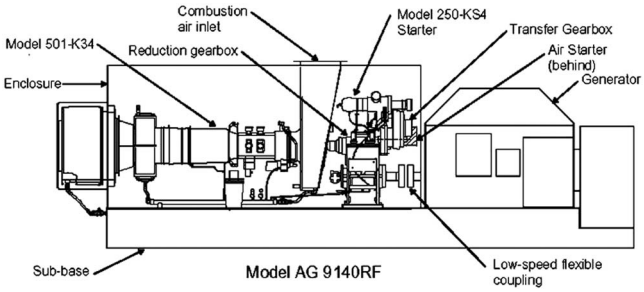


Fig. 1 SSGTG Model AG9140 RF

variety of reasons. The clutch assembly on all designs has been the primary failure mode due largely to entrained water in the air system, which results in periodic shock torque loading on the air turbine starter and associated clutch and then premature failure of clutch jaw teeth or pawls and ratchets, depending on the type of clutch used. Several attempts have been made to redesign the clutch and starter to make it more robust and able to withstand the extreme loading conditions that occur when water passes through the system.

When the DDG 51 SSGTG was developed, the U.S. Navy requested that the starter be redesigned to attempt to counter the failures experienced on previous ship classes. The redesigns of the starter to be more robust and to eliminate those failures of the clutch assembly were largely unsuccessful. Attempts were made by the shipbuilders to redesign the air systems to help with the removal of entrained water, but these attempts were largely unsuccessful due to the nature of shipboard air piping systems.

As with most programs associated with the marine gas turbine system, the U.S. Navy found that a large number of the successes came from removing primarily aeroderivative parts and replacing them with high-durability industrial components. The advantage of doing this is that the parts are often more robust and weight and size considerations are not as critical as those found on aircraft. In addition to incorporating industrial hardware, the U.S. Navy is very heavily driven by mean time between failure (MTBF) metrics and return on investment (ROI) calculations.

At this time, the ROI calculations have not been finalized, but the MTBF issues have been followed closely. The air turbine starter currently accounts for ~12% of the failures when analyzed by cost of parts per year, as shown in Fig. 4.

Air Turbine Starter and Starting System Development

The Parker (formerly Bendix) Model 36E-129 air turbine starter is an aeroderivative engine starter, and in conjunction with an air control valve Model 38E45-9A and a friction clutch assembly, it was designed to start the SSGTGs with Allison 501-K17 gas turbine on ship classes before the DDG-51 Class ships. The starter consisted of two sections: a turbine section with turbine wheel and planetary gearing, and a clutch section consisting of a Bendix-type

Ship	Starter Model	GT Model	GT Firing Speed	Maximum Engaging Speed	Clutch Overrunning Speed
DD963/DD993	Parker (Bendix) 36E-129	501-K17	2200 rpm	8400 rpm	13,820 rpm
CG-47	Parker (Bendix) 36E-152	501-K17	2200 rpm	8400 rpm	13,820 rpm
DDG-51	Allied Signal (Garret) ATS-100-459 & 467	501-K34	2200 rpm	9100 rpm	14,340 rpm

Fig. 2 U.S. Navy starter installations

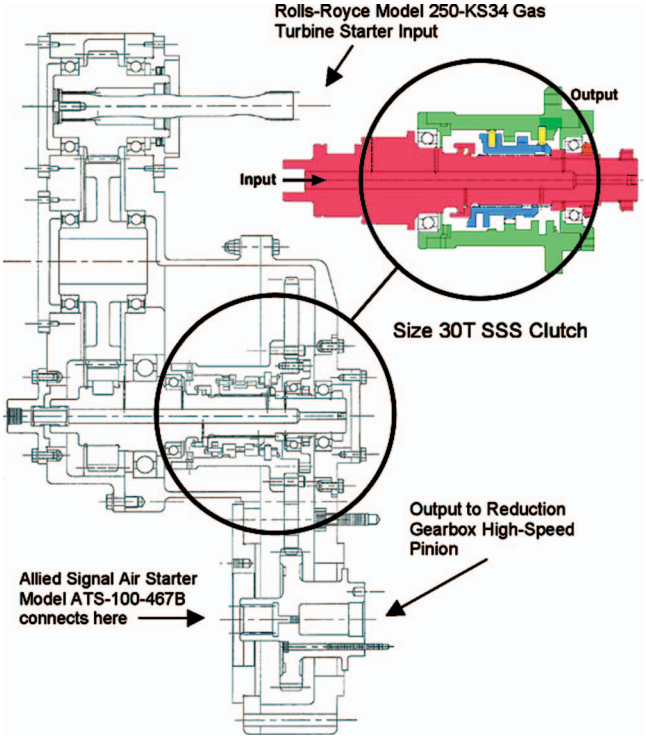


Fig. 3 RIMSS gearbox design

clutch with friction clutch to accommodate the inertial load at start-up and a jaw clutch with spring return to engage the air turbine reduction gear high-speed pinion.

The starter was designed for a normal inlet air pressure of 45 psig with rise rate of between 22 psi/s and 30 psi/s when the valve opened. For emergency operation, it was designed for 415 psig from the high-pressure air storage bottles with 147 psi/s rise rate, and the rise rate was closely controlled in both cases so that the jaw clutch would engage with minimum rotational angle so as to ensure repeated engagements without damage.

The instantaneous shock of engagement was absorbed by the friction clutch that incorporated a series of bevel washers and a lock nut to enable the clutch to slip at ~400 lb ft. When the air pressure was turned off and the gas turbine speed exceeded the starter speed, the shape of the jaw clutch teeth caused the jaw clutch on the turbine starter to move axially and disengage from the jaw clutch teeth on the component connected to the reduction gear pinion, and the return spring kept the starter jaw clutch disengaged.

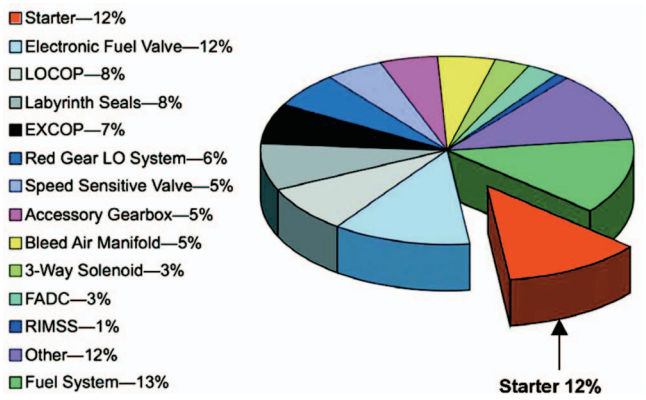


Fig. 4 Failure and parts cost analysis

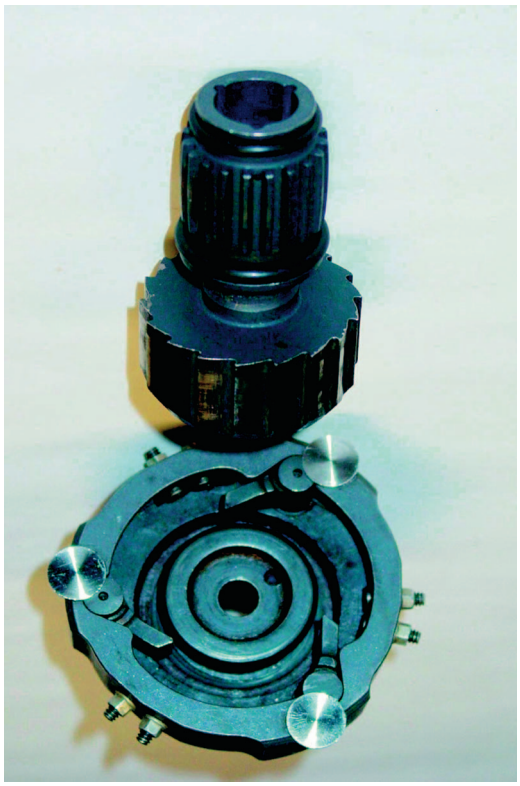


Fig. 5 Undamaged starter clutch

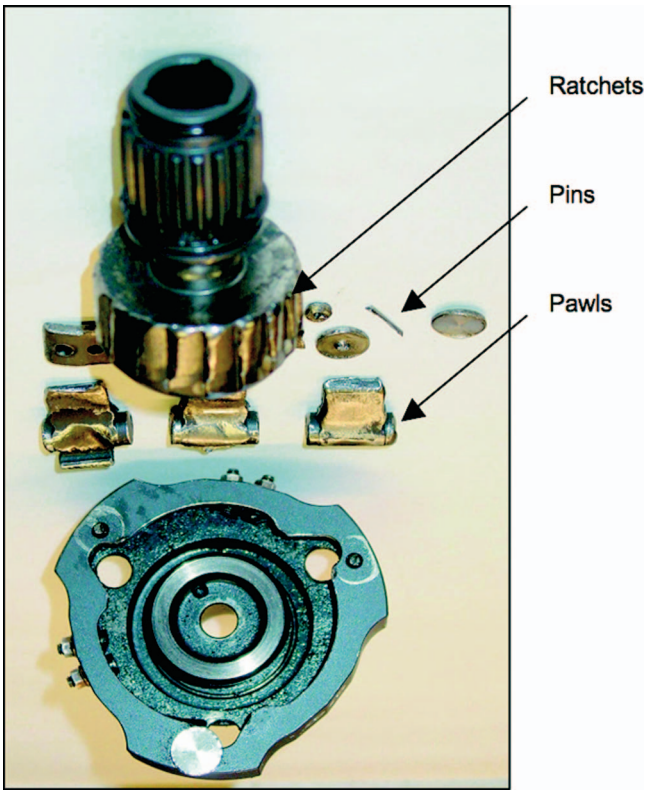


Fig. 6 Failed starter clutch

The Parker Model 36E-129 starters and associated equipment experienced problems and premature failures. As a result, the 36E-129 was phased out and replaced by the Model 36E-152. Design improvements were made to the planetary gearing, bearings, friction clutch, and jaw clutch to improve the design life by a factor of 2.3. However, it was found that the problems associated with water in the air systems and nonuniform loading of the starter resulted in continued failures of the jaw clutch. By the mid-1990s, nearly 20 starters every three months were failing, many with <200 cumulative starts. Primary cause of failures was related to shock torques from water in the air supply system.

When the DDG 51 SSGTGs were designed, the starter system was again modified to incorporate an alternate starter, a splined ratchet jaw connected to the starter output, and a pawl and ratchet clutch attached to the reduction gear pinion. Figure 5 shows the two halves of the starter clutch assembly. An Allied Signal ATS100-467 air turbine starter with $\sim$30% increased torque capacity and disengagement speed at 9100 rpm versus 8400 rpm was used. Though starter failures were reduced, the mode of failure remained the same.

When the AG9140RF SSGTG was introduced, the starter design was retained. However, the air turbine starter usage was reduced and an increased number of starts were done with the RRC Model 250 gas turbine to reduce failures associated with air turbine starts. Recently, the U.S. Navy found that problems associated with air turbine starts are still occurring and the failure of the pawl and ratchet clutch, as shown in Fig. 6, creates a cascading failure of the transfer gearbox resulting in the loss of redundant start capabilities. When the pawl and ratchet clutch components fail, the parts can become entrained in the transfer gearbox gear train, resulting in gear failure and disabling the ability to start with the RRC Model 250.

Design improvements within the SSGTGs that have been made by the U.S. Navy over the past few years have reduced the number and frequency of borescope operations necessary for the gas turbines and, hence, the number of times per year that the starter assembly has to be removed. At the time of this writing, the U.S. Navy is pursuing an SSS clutch that does not have the manual turning feature, as this will avoid geometry changes within the SSGTG module when the clutch is retrofitted and minimize the impact on overall installation cost.

SSS Clutch Basic Principle

The SSS Clutch is a freewheel-type, overrunning clutch that transmits torque through the full surface contact of surface-hardened gear teeth. Unlike the jaw clutch, this clutch engages and disengages automatically. The angle to engage is consistent on every engagement, being about 15 deg for starter clutch designs. Although not known precisely, the acceleration rates (revolutions per minute/second) of the starter motor for these U.S. Navy starter applications are well within the design capacity of the clutch.

On start-up, the clutch automatically engages the starter motor to the reduction gear pinion to drive the gas turbine and then disengages automatically when the gas turbine speed exceeds that of the starting motor. When disengaged with the clutch output at full speed and the starting engine at rest, there is no mechanical contact between the clutch input and output. If needed, a restart of the turbine can be done without waiting for the generator to be stationary.

The principle of operation of the SSS Clutch can be seen in Fig. 7. When the speeds of the clutch input and output reach synchronism, the pawls on one clutch element engage with ratchet teeth on the other to phase the teeth precisely for interengagement. Further shaft relative rotation causes the pawls to provide the small force to move the sliding component along helical splines, thereby engaging the driving and driven clutch teeth smoothly and positively.

The pawls do not transmit any driving torque because they axially move out of contact with the ratchet teeth before the clutch teeth have shifted into full driving engagement. The clutch completes its travel when the sliding component moves against a shaft

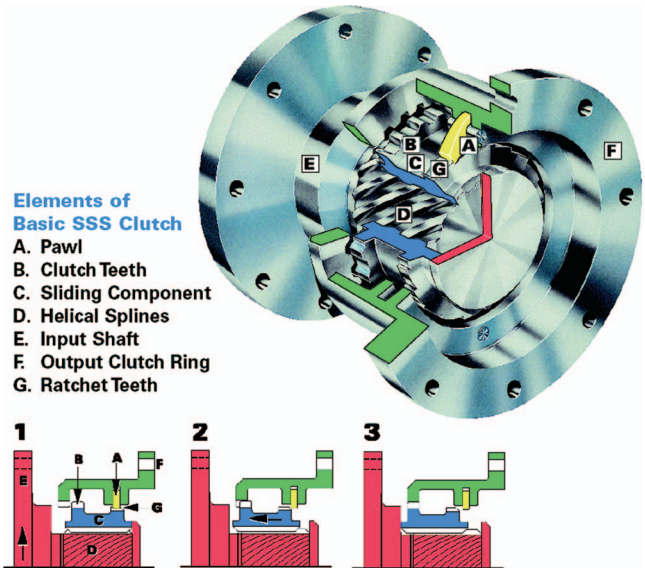


Fig. 7 SSS clutch principle of operation

abutment, and then, full torque passes through the helical splines and the fully engaged clutch teeth. As soon as the clutch input speed reduces relative to the clutch output speed, the clutch automatically disengages due to the reversal of torque on the helical splines.

To enable the clutch to engage at low and high speeds, but also to prevent sustained ratcheting action of the pawls with the clutch output at speed and the input stopped, primary and secondary pawls (PRISEC) are used, as can be seen in Fig. 8.

The primary pawls are mounted on the clutch sliding component, which, in turn, is mounted on the clutch output. These pawls are spring loaded into engagement with ratchet teeth in the bore of the input component. Unbalance relative to the pawl central pivot

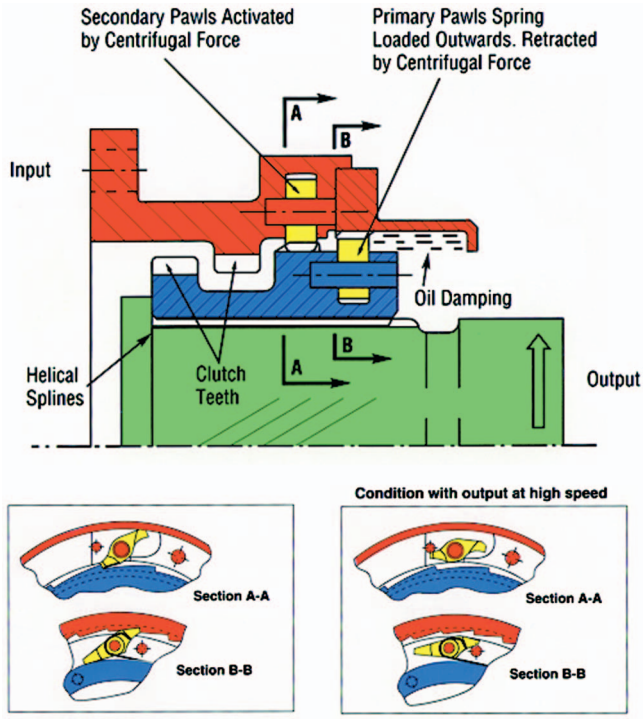


Fig. 8 Primary and secondary pawls (PRISEC)

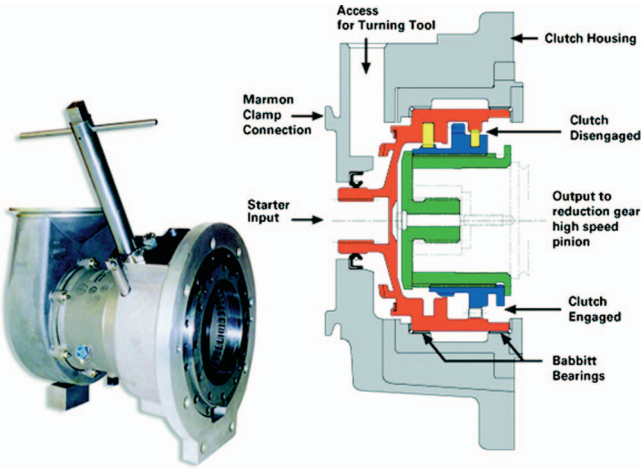


Fig. 9 Size 54T SSS clutch for DD 963 Class starters

pin enables the primary pawls to retract from the ratchet teeth due to centrifugal force when the output exceeds a predetermined speed, which is usually ~500 rpm.

The secondary pawls, which are mounted on the clutch input and engage with ratchet teeth on the sliding component, are used to engage the clutch in the high-speed range, which is to say when the clutch input accelerates to the same speed as the output. Unbalance relative to the pawl central pivot pin gives the pawl centrifugal movement toward the ratchet teeth when the input rotates. However, when there is high relative speed between the pawls and ratchet teeth, the pawl skims on the oil within the clutch until the differential speed between input and output is small.

With PRISEC pawls, both sets of pawls are inert when the clutch output is rotating at high speed and the clutch input is at rest or at low speed.

Initial SSS Starter Clutch Design

It was recognized that clutch axial length had to be kept to a minimum because there was not much axial space between the air inlet ducting and the SSGTG sound-proof module wall adjacent to the generator. This was important because any additional length of the SSS clutch versus the existing clutch would have caused interference. In order to keep the clutch length as short as possible and to utilize a proven design, a derivative of an SSS clutch first used in 1992 in the starting system of the General Electric MS-5001 gas turbine generator was chosen. Over 700 are now in service in MS-5001 and MS-6000 machines and are very reliable. Because the starting torque requirements were lower for the SS-GTGs, the clutch was redesigned to reduce the pitch circle diameter of the main torque transmitting teeth by ~1 in.

The new size 54 T SSS clutch was designed and produced in 1998. The clutch output sleeve was designed to be mounted on the external splines of the reduction gear pinion shaft extension and the clutch sliding component with main torque transmitting teeth to be mounted on the clutch output (see Fig. 9). Both were contained by an end plate bolted to the pinion shaft. The clutch input was designed to be supported and held radially and axially by an internal babbitt bearing in a stationary housing that was bolted to the SSGTG reduction gear housing using a one-piece adaptor plate. The clutch casing had a Marmon clamp interface for mounting the existing air turbine starter and a clutch input spline to connect directly to the starter shaft.

The SSS clutch is oil lubricated, whereas the jaw clutch needs no lubrication. The clutch housing was therefore designed with oil feed and drain piping that could connect into the existing reduction gear lubrication system. With only a 20 gpm total reduction gear lube system capacity, which could not be increased, the clutch had to be designed for minimum oil flow.

To meet the requirement of manually turning the high-speed pinion and gas turbine, a bevel gear was incorporated into the clutch input component and the housing was designed to radially accept a cylindrical turning tool with bevel pinion at the end. A small cover could be closed and locked on the clutch housing when the turning tool was removed.

Size 54T SSS Clutch Testing at NSWCCD

Before the clutch could be installed onboard a U.S. Navy ship, qualification testing was necessary on a land-based engineering site (LBES) at the Naval Surface Warfare Center, Carderock Division, (NSWCCD), Philadelphia. Two existing air turbine starters were modified to be able to connect the starters and new clutch assemblies, as shown in the left view of Fig. 9. The starter assemblies were installed on the LBES in late 1998 to do 250 clutch engagements consisting of 200 cycles to "motor" without firing the gas turbine and 50 turbine starts. This test represented approximately two years' shipboard operation for the starters and clutch assemblies.

In accordance with a preapproved test procedure, the clutch was to be subjected to 200 "motors," whereby the gas turbine generator was accelerated to 3000 rpm and then allowed to decelerate to rest. Duration of each "motor" was between 15 s and 2 min. When the valve shut off air to the air turbine starter, it decelerated to rest in 5 min, whereas the turbine generator took 3–4 min to stop.

Testing took place over several months, was completed in 1999, and proved the SSS clutch suitable for the application. Minor problems occurred with the manual turning tool in the aluminum clutch housing and with maintaining the tool's bevel teeth in mesh with the teeth on the clutch input assembly. A design change to enable the tool to be retained in the housing when in use and to use needle roller bearings to avoid direct contact with the housing was adopted, and the problems were solved. Failure of an air turbine starter motor in August 1999 ended the test prematurely, with less than ten motoring cycles left to complete.

Because of increasing problems with DDG 51 air turbine starter clutches, priority was shifted to the design, testing, and installation of another SSS clutch design for those SSGTGs.

DDG 51 Starter Systems and Test at NSWCCD

The DDG 51 starter system was originally designed to use an Allied Signal (Garrett) Model ATS100-467B air turbine starter with pawl and ratchet clutch to start the RRC Model AG9130 SSGTG with RRC Model 501 K-34 gas turbine. A slightly smaller size 45T SSS clutch was proposed that was similar to what had been previously tested at NSWCCD, Philadelphia, but with design improvements incorporated (see Fig. 10).

The design torque requirements for the DDG 51 SSGTG starter clutch were 285 lb ft (standard day stall torque), 326 lb ft maximum impact torque, and 600 lb ft ultimate torque. The original pawl and ratchet clutch incorporated a shear section to prevent overtorque damage, but it was not desirable to incorporate such a feature into the new SSS clutch. Two prototype SSS clutches were produced in 2000 and delivered to NSWCCD, Philadelphia, for installation on the SSGTG LBES.

The prototype 45T SSS Clutch was designed to be installed on a Cincinnati Gear Company reduction gear, but the LBES had a WesTech Gear Company gearbox. After suitable adaptors were manufactured, the size 45T SSS clutch was installed. A modified air inlet duct "banjo" was produced to reduce the overall axial length of this new starter assembly so that it would fit into the SSGTG module for test. The clearance between banjo and interior module wall was 1–1/8 in when installed.

Endurance testing of this new assembly started in midyear 2000 and consisted of 250 total starts. Two hundred test starts were actually motors where the air turbine starter held the drive train at about 4000 rpm for 30–60 s without firing the gas turbine, and 50 were true turbine starts, with the gas turbine operated at full speed for 3–5 min.

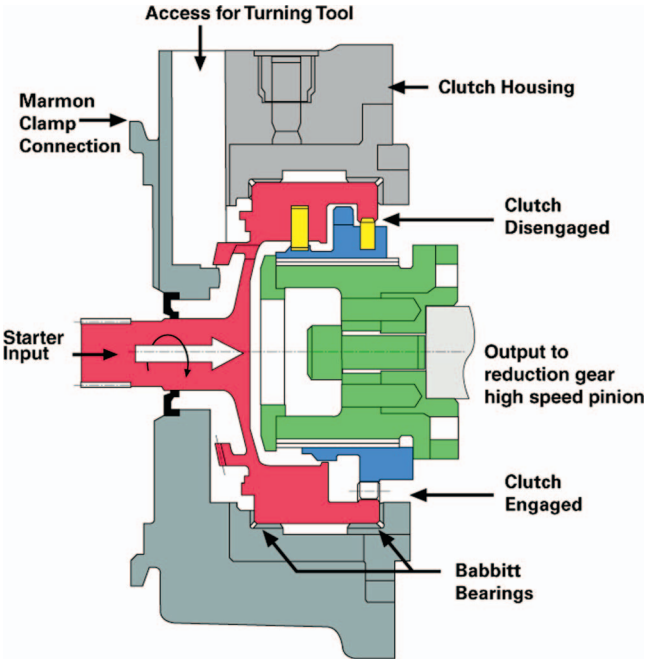


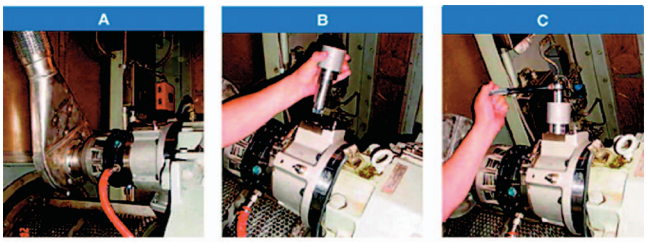
Fig. 10 Size 45T SSS clutch for DDG 51 Class starters

During these motors, a prelube oil pump was started for 15 s and the oil pressure was about 15 psi, with oil flow at 1.5 gpm. After the pre-lube pump stopped, the oil pressure reduced to about 4–5 psi from the shaft-driven pump when the drive train was ~4000 rpm. The oil flow at 4000 rpm was ~1.5 gpm, which was satisfactory. The initial "motors" confirmed the clutch engaged and disengaged satisfactorily, and at 9000 rpm the oil flow to the clutch was 2.4 gpm, which was acceptable.

When the air turbine starter was shut down, it did not initially come to rest with the output at 14,400 rpm, due to oil retained within the clutch assembly. Two, 1.5 in. (1.25 in. i.d.) drains were added to the clutch housing, and that solved the drag issue. Subsequently, the air turbine starter repeatedly came to rest with the gas turbine operating at full speed. The torque to manually turn the gas turbine generator was measured on an SSGTG on the USS Porter, DDG 78, and a maximum torque of 150 lb ft was recorded. This contrasted with 75 lb ft measured on the LBES. The tool's bevel gear was increased to a 4:1 ratio and subsequent testing demonstrated the turning tool capability of 44 lb ft, or about 175 lb ft at the starter pad.

The LBES testing was completed in 2001, and two prototype size 45T SSS clutches were installed in June 2002 on SSGTGs on the USS Ramage, DDG 61, in order to gain further operating experience. Figure 11 shows the installation, the way the manual turning tool is attached to the clutch housing, and how the manual turning tool is used with a standard socket wrench and ratchet assembly.

These two size 45T SSS clutches have operated without problems for approximately four years, with the exception of one minor oil leak, which occurred due to the removal of a starter and reinstallation without a clutch input shaft o-ring. In this period of time, these starting clutches have had the following operating history (see Fig. 12), and the turning tool has been used approximately every three months to enable borescopic inspection of the gas turbine.



A Size 45T SSS Clutch installed onboard DDG-61, USS Ramage.
B Locked access plate removed and turning tool being inserted into clutch.
C Tool rotating gas turbine for borescopic inspection.

Fig. 11 Installation onboard USS Ramage

DDG 51 Starter System on Ships With RIMSS

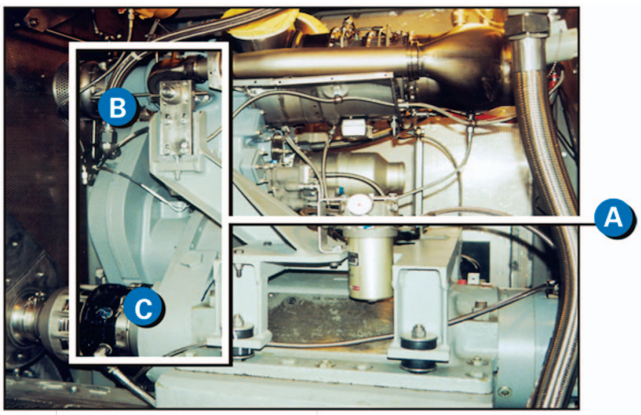
U.S. Navy ships DDG 51 through DDG 67 were constructed with the RRC Model AG9130 SSGTG, which relied on high-pressure air bottles for a dark-ship start. For DDG 68 through DDG 78, the AG9130 SSGTG was redesigned for greater overall maintainability and reliability, and designated AG9140.

DDG 79 through DDG 112 utilize SSGTG Model AG9140RF, incorporating the RIMSS and RRC Model 250-KS4 gas turbine starter, as shown in Fig. 13, to provide redundant starting capability and dark-ship start capability as the HP air system was eliminated. The AG9140RF SSGTGs use the size 30T SSS clutch in the RIMSS transfer gearbox, as shown in Fig. 3, to automatically disconnect the Model 250-KS4 gas turbine.

The air turbine starter interface for ships with RIMSS is different from those without RIMSS, and it is not possible to utilize the size 45T SSS clutch on these ships. However, the RIMSS gear casing has sufficient space to enable a size 38T SSS clutch to be installed inside the gear casing, and the clutch output component can fit inside the bore of the RIMSS final gear and connect to the high-speed pinion of the reduction gear.

SSGTG	Operating Hours	Starts
No. 1 SSGTG	5078.5	374
No. 2 SSGTG	6876.6	340

Fig. 12 Experience onboard USS Ramage



A	RIMSS Transfer Gearbox Design (see FIGURE 3)
B	Size 30T SSS Clutch (see FIGURE 3)
C	Size 38T SSS Clutch (see FIGURE 14)

Fig. 13 RIMSS transfer gearbox installation

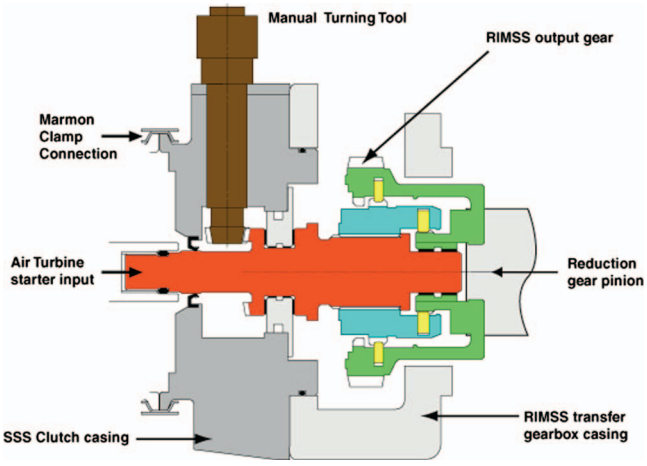


Fig. 14 Size 38T SSS clutch with turning tool

The SSS clutch therefore has been designed to mount into the RIMSS transfer gearbox at the air turbine starter end to replace the pawl and ratchet clutch, as shown in Fig. 14. A manual turning tool has been included on the clutch's short input shaft to enable borescoping the RRC Model 501 gas turbine without having to remove the air turbine starter assembly.

The cost of the Size 38T SSS clutch with manual turning capability is about three times that of the pawl and ratchet clutch. Because of its length, modifications would be needed to the air inlet duct banjo for the air turbine starter to accommodate an SSS clutch retrofit and enable the assembly to be installed in the existing module. This modification work would further increase the installation cost.

A U.S. Navy ROI study has shown that, with increased reliability of the RRC Model 501 gas turbines and hence, reduced frequency of borescope operations, the total increased installation cost associated with the clutch manual turning feature and modified air inlet duct banjo is not justified. An SSS clutch design that does not incorporate the manual turning tool is shown in Fig. 15. This design will allow for installation into the SSGTG module without modification to the air inlet duct banjo, greatly reducing the total installation cost.

The pawl and ratchet clutches periodically fail, and failure of the pawl and ratchet clutch causes costly collateral damage within the RIMSS gearbox. With the installation of the size 38T SSS clutch with proven robust design and reliability, the reliability of the starting system of AG9140RF would be improved. If the return on investment (ROI) proves to justify the manual turning tool device for borescoping the RRC Model 501 gas turbine, then the U.S. Navy may consider this option.

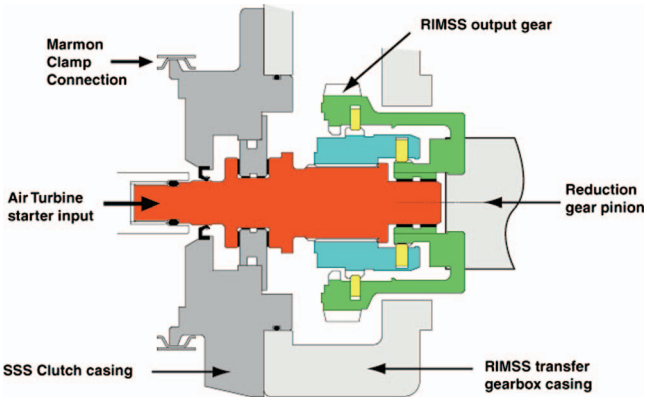


Fig. 15 Size 38T SSS clutch without turning tool

Conclusion

These SSS clutch starter designs are based on more than 40 years of application experience in industrial gas turbine starting drive systems as well as thousands of gas turbine main drive disconnect applications in power generation and marine propulsion systems worldwide. The design and development of these clutches for U.S. Navy gas turbine generator air turbine starters show that a proven design for industrial starter applications can be reliably applied to U.S. Navy shipboard starters with appropriate interface changes.

The value of testing on a land-based engineering site before installing onboard a ship has been demonstrated, as the two size 45T SSS clutches retrofitted on the USS Ramage have operated without problems for nearly four years. Nearly 100 size 30T SSS clutches have been supplied for RIMSS transfer gearboxes for dark-ship starts, and more than half are in service and have had trouble-free operation.

The size 38T SSS clutch design can be integrated into the RIMSS transfer gearbox to provide a more reliable clutch for air turbine starts for DDG 51 Class gas turbine generating sets. It may be possible to apply the size 38T SSS clutch in ships without RIMSS, but the size 45T SSS clutch results in slightly shorter installed length, which may be advantageous to reduce overall installation costs.

The manual turning tool was tested and found to be a potentially useful tool to reduce time to borescope the Model 501 gas turbines and potentially increase the installation reliability because the starter assembly, once installed, would not have to be removed except for maintenance of starter components themselves. The current U.S. Navy philosophy used to incorporate improvements to shipboard equipment is heavily driven by ROI calculations, and as a result, it remains to be determined whether this additional feature will remain in the design after the U.S. Navy completes all ROI calculations.

Further prototype design efforts and fleet evaluations will be required to provide endurance, MTBF and maintainability metrics, and the applicability of the design in the AG9140RF SSGTGs. Additionally, ROI calculations will be required by the U.S. Navy to warrant proposed changes detailed in this paper [1–6].

Acknowledgment

The authors acknowledge the Naval Surface Warfare Center, Carderock Division (NSWCCD), Gas Turbine Code 935, Philadelphia, PA, for its support and technical contribution in development and testing of these synchronous self-shifting starting clutches for U.S. Navy gas turbine generating sets.

References

[1] Skladony, R. M., and Hendry, M. L., 1993, "Starting System Upgrade for General Electric Frame 5 Gas Turbine Generators," ASME Paper No. 93-GT-181.
[2] Masincup, W. E., Halsey, J. E., and Russom, D., 1996, "Redesigning a Generator Set for Improved Maintenance," ASME Paper No. 96-GT-215.
[3] Patterson, J. S., 1997, "Allison 501-K17 SSGTGS Technical Directive Experience," ASME Paper No. 97-GT-169.
[4] Masincup, W. E., and Jackson, S., 1997, "Alternative Starting Methods for Shipboard Gas Turbine Generators," ASME Paper No. 97-GT-251.
[5] Overton, Jr., L. L., Masincup, W. E., and Halsey, J. E., 1997, "A Mechanical Start System for U. S. Navy Destroyer Generator Sets," ASME Paper No. 97-GT-252.
[6] The Bendix Corporation, 1974, "Technical Manual—Installation, Operation, Maintenance, and Repair Instructions with Parts List. Air Turbine Starter Assembly Part No. 36E129-2A," NAVSEA 0941-LP-049-4010, pp. 3-1–8-2.

Design and Off-Design Analysis of a MW Hybrid System Based on Rolls-Royce Integrated Planar Solid Oxide Fuel Cells

Loredana Magistri
Thermochemical Power Group,
Department of Energy Systems,
University of Genoa,
Genoa, Italy

Michele Bozzolo

Olivier Tarnowski

Gerry Agnew

Fuel Cell Systems,
Rolls-Royce,
Derby, UK

Aristide F. Massardo
Thermochemical Power Group,
Department of Energy Systems,
University of Genoa,
Genoa, Italy

In this paper the design point definition of a pressurised hybrid system based on the Rolls-Royce Integrated Planar-Solid Oxide Fuel Cells (IP-SOFCs) is presented and discussed. The hybrid system size is about 2 MWe and the design point analysis has been carried out using two different IP-SOFC models developed by Thermochemical Power Group (TPG) at the University of Genoa: (i) a generic one, where the transport and balance equations of the mass, energy and electrical charges are solved in a lumped volume at constant temperature; (ii) a detailed model where all the equations are solved in a finite difference approach inside the single cell. The first model has been used to define the hybrid system lay out and the characteristics of the main devices of the plant such as the recuperator, the compressor, the expander, etc. The second model has been used to verify the design point defined in the previous step, taking into account that the stack internal temperature behavior are now available and must be carefully considered. Apt modifications of the preliminary design point have been suggested using the detailed IP-SOFC system to obtain a feasible solution. In the second part of the paper some off-design performance of the Hybrid System carried out using detailed SOFC model are presented and discussed. In particular the influence of ambient conditions is shown, together with the possible part load operations at fixed and variable gas turbine speed. Some considerations on the compressor surge margin modification are reported.
[DOI: 10.1115/1.1839917]

1 Introduction

Many papers have been published in the last few years concerning the analysis of Hybrid Systems (HS) based on Solid Oxide Fuel Cells (SOFCs) technology [1–6]. These works are of great interest since they have pointed out the very high efficiency of SOFC hybrids and some new aspects that must be carefully investigated for their full development [7,8].

Unfortunately, most of the mentioned investigations suffer from a lack of information about the characteristics of the HS components such as fuel cell stack, the turbomachinery, the heat exchangers, etc. In fact, in many published papers simplified hypothesis, particularly about cell geometry, compressor maps, and so forth have been often used to carry out the hybrids design point and off design analyses.

Another critical aspect of hybrids investigation is related to the model used for the cell-stack analysis. In fact in many published papers the simulation has been carried out using very simple models (i.e., correlation) or using lumped volume approach (0D). These approaches are very useful for preliminary HS investigations but they cannot give a complete description of the cell internal thermochemical behavior to avoid unfeasible cell and HS operating conditions as already shown by the authors [7]. More complex models are necessary, but on the other hand to generate a very detailed model the cell geometry and the electrochemical characteristics must be known in a good detail. This information *is* usually not available in open literature. Moreover, to develop a reliable cell model it is necessary to validate theoretical results with apt experimental data (in a similar way as normally done in the field of turbomachinery with the so called Test Cases).

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Atlanta, GA, June 16–19, 2003, Paper No. 2003-GT-38220. Manuscript received by IGTI, October 2002, final revision, March 2003. Associate Editor: H. R. Simmons.

In order to overcome all the barriers previously discussed the investigation of hybrid systems based on the Rolls Royce Integrated Planar IP-SOFC technology has been carried out in collaboration with stack developer: Rolls-Royce plc, turbomachinery manufacturers and model developer: Thermochemical Power Group—TPG—at the University of Genoa. This work is part of the EC funded Framework 5 project IM-SOFC-GT co-ordinated by Rolls Royce. In IM-SOFC-GT different turbomachinery manufacturers are in charge of different FC/GT hybrid power sizes (Turbec: subMWe, Turbomeca: 1–2 MWe, ABB Turbo Systems, Rolls Royce, and Alstom Power Sweden: 20–30 MWe). In this paper the size of the investigated system is just under 2 MWe. Therefore, a Turbomeca gas turbine has been considered.

The analysis of the 2 MWe size IP-SOFC-HS has followed five main steps: (i) development of an IP-SOFC-HS generic model used to define the pressurised HS lay out and the design point conditions. (ii) Development of an IP-SOFC detailed model and its validation using Rolls Royce experimental data. (iii) Verification of the HS design point using the detailed model and suggestion of modifications to meet all the system defined constraints. (iv) Analysis of the ambient conditions influence on the HS performance. (v) HS part load analysis at fixed and variable turbine speed, taking into account IP-SOFC, compressor surge margin, turbine, and heat exchanger constraints.

The 2 MWe HS lay out, its on and off design performances are presented and analyzed in details. Strong cooperation between fuel cell stack developer Rolls-Royce, turbomachinery manufacturers, and TPG at the University of Genoa which have expertise in different areas: electrochemical, turbomachinery, energy systems, etc., has been essential to develop a full reliable model of the Rolls Royce IP-SOFC-HS.

The results for sub-MWe, using a Turbec micro gas turbine, and multi-MW system (20–30 MWe) with an Alstom gas turbine, will be presented and discussed in a future paper.

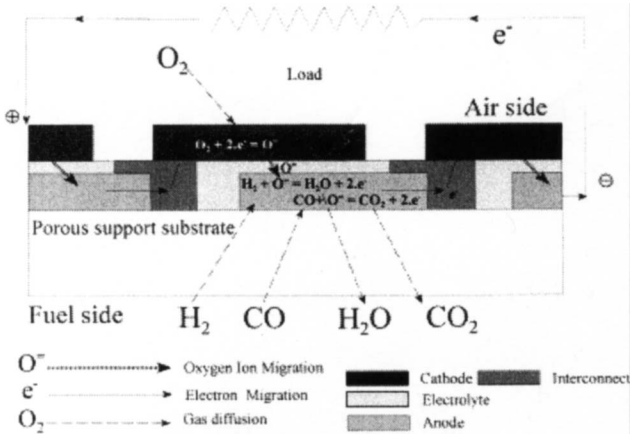


Fig. 1 Multi cell MEA concept

2 Integrated Planar SOFC Model

The SOFC stack considered in the present work is the Rolls-Royce Integrated Planar Solid Oxide Fuel Cell stack. IP-SOFC is a power system based on the "multi-cell membrane electrode assembly (multi-cell MEA) module" [9] carrying many series connected cells with supported electrolyte membranes only 10–20 μm thick (Fig. 1). Every module produces about 50 We. 10 modules form a bundle. A strip is made of 6 bundles producing 3 kWe. Finally, 16 strips form a 50 kWe power block with an overall power density of about 200 mW/cm^2 (see Fig. 2). At atmospheric running conditions nominal stack efficiency is about 50% [10].

In order to simulate the performance of the SOFC stack, both as a stand alone unit and in the Hybrid System configuration, different fuel cell models have been developed by TPG. Two different approaches have been used to understand how discriminating the accuracy of the simulation is in HS modeling. The first model, named "*Generic model*," is based on the overall mass and energy balances coupled with appropriate expressions for the reaction kinetics, thermodynamic constants and material properties. The balance equations have been solved between the inlet and outlet of each component of the stack, evaluating average values for physico-chemical variables and estimating the electrochemical performance of the stack.

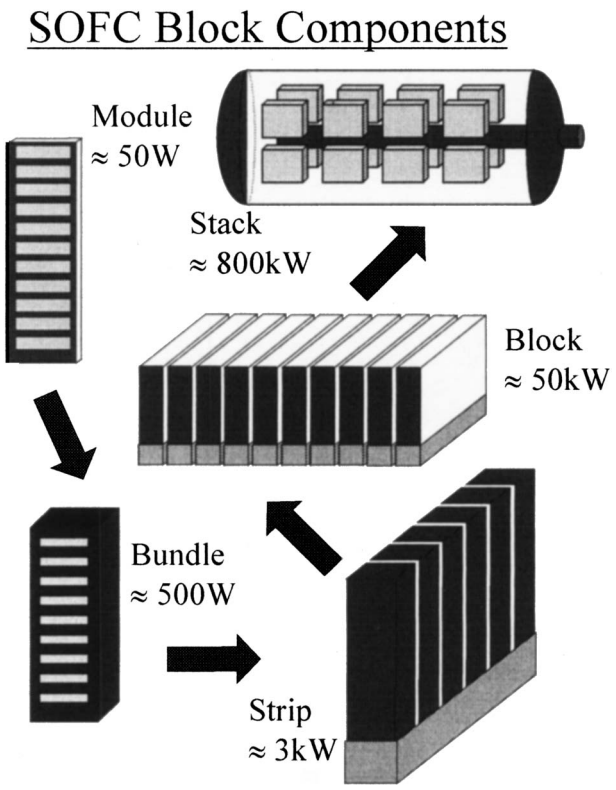


Fig. 2 SOFC block components

The following main assumptions have been made: (i) adiabatic reformer (ii) reforming and shift reactions at equilibrium in the reformer and in the SOFC stack.

In the stack, the relation current–voltage is worked out by evaluating the thermodynamic voltage.

The current–voltage behavior of the stack is evaluated by subtracting the voltage losses from the thermodynamic potential corresponding to the value of the electrical current. This SOFC model is similar to one developed previously and described in detail in [10]. The SOFC model has been developed using MATLAB-Simulink language. Parametric and off-design studies have been carried out and the results have been compared to experimental data with good agreement, both for atmospheric and pressurised conditions.

A second model has been developed with the aim to obtain a better understanding of the thermochemical characteristics behavior inside the cells. However, taking into account that one of the main goal of this model is to be integrated in the HS whole code the approach must be as complete as possible (detailed information of the internal cell characteristics distribution), and in the same time very easy to be managed at HS simulation level. Therefore, a one-dimensional approach has been used [8]. The model is based upon the following assumptions: (i) one-dimensional model along the x-coordinate, along the anodic path; (ii) uniform distribution of the anodic flow rate among the inner channels of the module; (iii) all the gases flow in the axial direction, no transversal flows are considered; (iv) all the cells of the module supply the same electrical current; (v) mass and energy balances are written in the plug-flow form along the axial coordinate, (vi) no heat exchange by radiation between gases and solid; (vii) air is considered as the cathodic reactant and a mixture of $H_2/H_2O/CO/CO_2$ is considered as the anodic one; and (ix) shifting reaction is assumed at the equilibrium in the anodic compartment. The cell model includes the evaluation of the electrochemical performance. The current–voltage behavior of the stack is evaluated by subtracting from the thermodynamic potential (i.e., the voltage of the stack under conditions of thermodynamic reversibility, the overall voltage losses. In this model Ohmic losses and activation losses, due to sluggish electrode kinetics have been simulated; moreover, concentration losses occurring when the diffusion of the reactants through the electrodes is slower than the electrochemical reaction, have been taken into consideration. In this way it is possible to evaluate the electrical power supplied by the cell. The evaluation of the electrical performance is made locally along the stack length: indeed, in order to integrate the model, the stack is divided into a number of grid points (corresponding to a single cell), and the evaluation of the electrical performance is repeated in each of them. The differential equations of the model have been integrated using a relaxation method for the Fourier problem and the charge balance, combined with a finite difference method to solve the other equations. Input data are geometrical characteristics, operating conditions, inlet flow conditions and gas and material properties [8].

With this model is possible to evaluate the single cell performance and the behavior of the overall stack. However, the passage from the single cell to the stack size cannot simply be made by assuming that all the cells of the stack show identical behavior to the single cell operated under adiabatic conditions. Indeed, in the stack the effects of radiation between contiguous cells cannot be neglected, which tend to smooth the temperature profile along

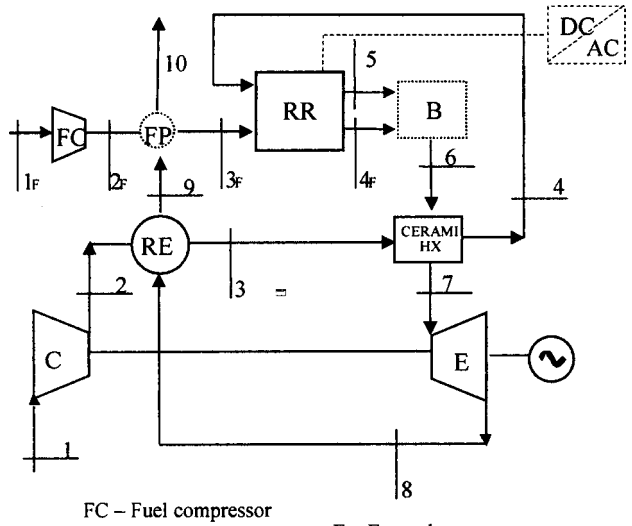


Fig. 3 Hybrid system simplified lay-out

each cell. The fuel cell stack includes also a nonadiabatic reformer unit that has been simulated with the same code in both the different fuel cell models.

The model results have been compared to the experimental data obtained from both a small scale IP-SOFC module and a full-size prototype; in both cases the agreement is good [8].

3 Hybrid System Configuration

In a previous paper IP-SOFC hybrid system layouts, performance and technological constraints have been presented both for atmospheric and pressurised cycles using a generic SOFC model [11]. The main result of the mentioned work has been the evident superiority of the pressurized configuration versus atmospheric one, in terms of efficiency, specific power, and cost of electricity. Moreover, the technological constraints are less stringent in the pressurised case, while in the atmospheric case the constraints on high temperature heat exchanger are very difficult to comply with.

From the economic point of view the pressurised solution seems to be more effective taking into account that, at fixed fuel cell size, the generated power increases and therefore the specific power cost reduces [11,12]. For these reasons this work focuses only on the pressurised cycle solution.

Figure 3 shows a simplified lay out defined for a 2 MWe power system. The stack outlet gas stream feeds the turbine at a temperature lower than 950°C (to verify the GT maximum temperature—uncooled GT—and life constraints), and provides power both to the compressor and the electrical load. In the cited figure both cathodic (air) and anodic (fuel) compressors and preheaters, in order to satisfy the inlet requirements of the fuel cell stack, are shown. An interesting feature of this solution is that it is not necessary to provide the system with a high efficiency high temperature heat exchanger (Ceramic Heat Exchanger), in fact a value lower than 50%, compatible with the current technology, can be used. It is also possible to provide the plant with a District Heating system to recover exhaust heat downstream the recuperator. To maintain the excellent environmental characteristics of the HS no afterburner has been considered in this work; however it is possible to use it just to have a proper value for the TIT. All the component characteristics have been defined between TPG, Rolls Royce and Turbomeca using as a tool the generic IP-SOFC-HS model.

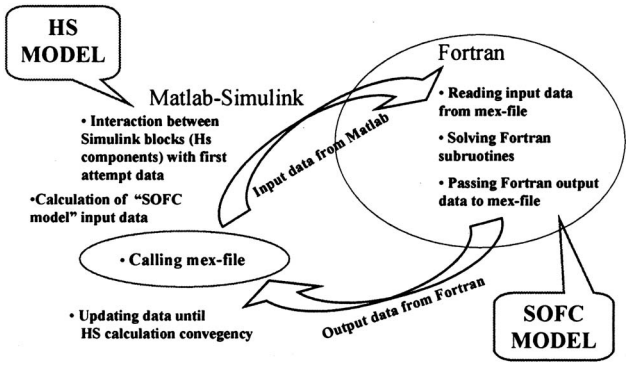


Fig. 4 Code structure

4 Design Point Definition

The first step of the work has been the definition of the design point characteristic. Therefore, for the design point definition of the Hybrid System, and for the following off design HS studies, apt models have been created combining the FC Stack models (generic or detailed) with a whole system model where all the components, such as compressors, expanders, heat exchangers and electronic devices, have been modelled in detail [10]. The HS component models and the generic IP-SOFC model are written in MATLAB-Simulink language thus the interaction is very easy to be organized. Since the detailed IP-SOFC model is written in Fortran language it is necessary, as shown in Fig. 4, to create an apt interface between the two different languages through an appropriate function (mex-file).

The Design point of the HS plant has been chosen among the operating conditions where turbine flow rate, rotational speed, pressure, turbine inlet temperature are compatible with the limitations of this technology and with the size of the machine considered by Turbomeca.

The design point conditions evaluated with the two IP-SOFC models are reported in Table 1. The generic model has been used at the beginning to provide a first attempt set of data for design point calculations. Then, using this data the preliminary design of a gas turbine (geometry and maps) has been carried out, and the performance of the IP-SOFC-HS has been evaluated again using the detailed cell-hybrid system model.

In the first calculation (generic SOFC model) the fuel cell is fed with the same air flow rate as the gas turbine, operates at the GT pressure, and the stack exhausts temperature has to match the turbine inlet temperature (TIT) required by the GT design point operation. In this case the fuel cell size has been scaled up considering the ratio between the power requested for this HS size and the power supplied by a single cell unit. Due to the generic SOFC model hypothesis the outlet temperatures of the cathodic and anodic streams have been considered equal to the average stack temperature; in these conditions design point operating tem-

Table 1 Design point operating conditions and performance

	Simplified SOFC Model	Detailed SOFC Model	Detailed SOFC Model
Air Flow Rate	2.731 [kg/s]	2.731 [kg/s]	3 [kg/s]
Fuel Flow Rate	0.06 [kg/s]	0.06 [kg/s]	0.06 [kg/s]
Compr. Pres. Ratio	6	6	6
Stack Temperature	900 [°C]	See Fig. 5(*a*)	See Fig. 5(*b*)
Fuel Cell Power	1608 [kW]	1688 [kW]	1631 [kW]
Fuel Cell Efficiency	0.536	0.562	0.54
Turbine Inlet Temp.	912 [°C]	807 [°C]	755 [°C]
HS Power	2040 [kW]	1986 [kW]	1914 [kW]
HS Efficiency	0.68	0.662	0.63

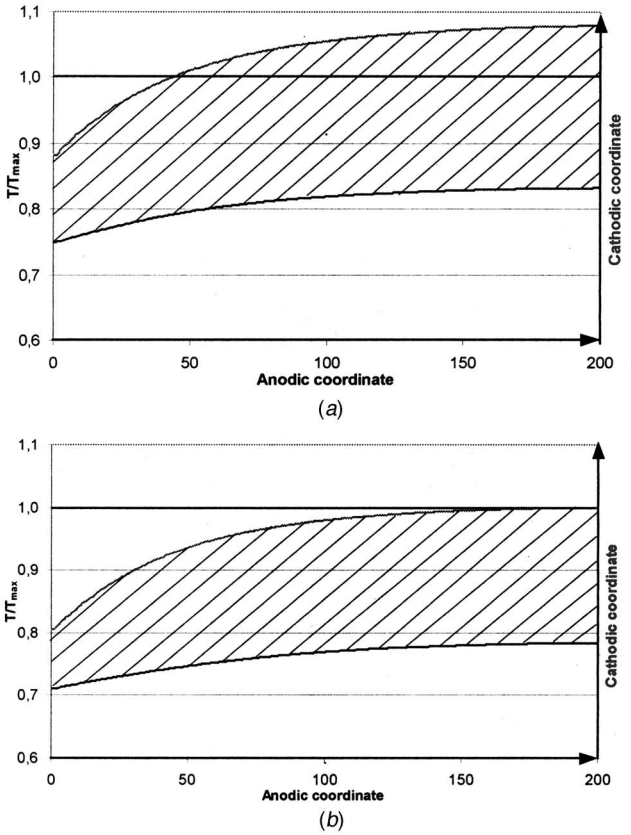


Fig. 5 (a) Temperature distribution inside the stack. (b) Temperature distribution inside the stack.

perature is 900°C. Turbine inlet temperature is 912°C taking into account the combustion of the anodic exhaust flow and the heat exchange with the air flow after the recuperator.

The HS design point calculation has been carried out also using the detailed IP-SOFC model in order to investigate the behavior of thermochemical cell parameters and the feasibility of the design conditions assumed using the generic model. In this second case the fuel cell stack has been scaled up considering a finite number of strips supplying the opportune power for the system (this is the first constraint to be considered using the detailed model since the number of cells is correlated to the number of bundles). In this case the overall stack and HS performance are in good agreement with the results of the previous calculation. However, the analysis of temperature distribution inside the stack, shown in a nondimensional form in Fig. 5(*a*), highlights several zones where the operating temperature is higher than a Rolls Royce fixed technological limit. These temperature peaks could lead to deterioration of the cell components (especially the anode, where Ni is present) or even breakages within the stack. The simplified model, that estimates only the average value of the stack temperature, did not allow to study these temperature peaks, and thus the plant design might lead to the choice of design point conditions detrimental to the SOFC group operation (unfeasible solution).

In order to avoid these problems it is necessary to decrease the maximum cell temperatures. A solution may be an increase of air mass flow rate keeping constant the current density and the fuel utilization factor at the defined IP-SOFC design point value. This choice directly affects the design point mass flow rate of the gas turbine coupled with the fuel cell. In this case also the size of the gas turbine must be revised to have a proper GT design point consistent with the new HS conditions.

Figure 5(*b*) shows that by increasing the air flow, at fixed pressure ratio, all the operating envelope for the stack is kept in a safe zone, avoiding cell temperature above the fixed limit (feasible design point).

The design point obtained in this case and the performance of the plant are also shown in Table 1.

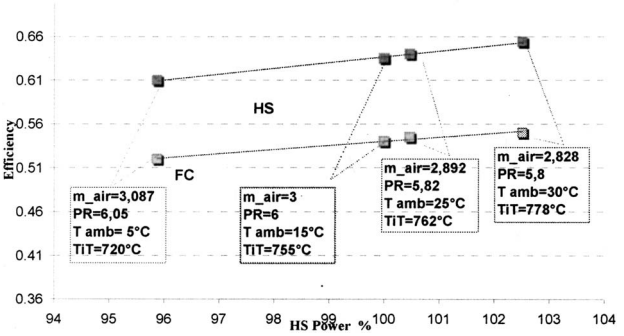


Fig. 6 Influence of ambient temperature without control system

5 Off Design Analysis

When the design point conditions are defined another step of the HS investigation is the off-design analysis. This aspect may be easily carried out with the models developed by TPG with Matlab-Simulink as already shown in previous works for different cell technology. Matlab-Simulink will also be very useful for future transient investigations.

In this paper two different off-design conditions are presented: (i) ambient conditions influence; (ii) part-load conditions.

To carry out the off-design analysis complete maps of about 400 kW sized gas turbine compressor and expander have been supplied by Turbomeca to TPG. At off-design conditions the operating point of the HS on the compressor map has been carefully defined including the evaluation of the actual surge margin. This way, only safe operating conditions are considered.

5.1 Ambient Conditions Influence. Figure 6 describes the influence of ambient temperature (in the range 5–30°C, design point at 15°C) on the FC stack operation and on the whole hybrid system. No modification has been considered for cell fuel flow rate and current density (equal to the design point values at $T_{amb}=15°C$).

Operating temperatures of the plant are strongly affected by ambient conditions. With an increase in compressor inlet temperature, all temperatures in the cycle increase; this way the stack operates in more favorable electrochemical conditions, with lower overpotentials, producing more power; as a direct consequence the whole system power output increases, together with both electrical efficiencies (FC and HS). But in some cases the IP-SOFC temperature could be too high and some problems could occur (Fig. 7) as already discussed for the design point definition [see Fig. 5(*a*)]. In fact according to the mentioned figure it is possible to note that an increase of about 15°C (from 15°C to 30°C) corresponds to an increase of about 4% of the maximum IP-SOFC temperature.

To avoid these problems it is necessary to modify the cell operating point varying both the fuel flow rate and the current density keeping the stack temperature as close as possible to the design point conditions. In this case the behavior of the plant is very different to the previous one. Considering Fig. 8 it is possible to observe that an increase of the ambient temperature leads to a decrease of the power supplied by the IP-SOFC and by the gas turbine, as usually happens in the traditional GT plants. But in this case the system efficiency increases due to the better performance

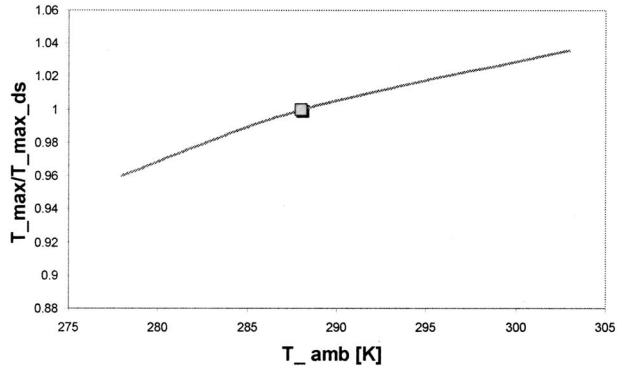


Fig. 7 Influence of ambient temperature on FC temperature

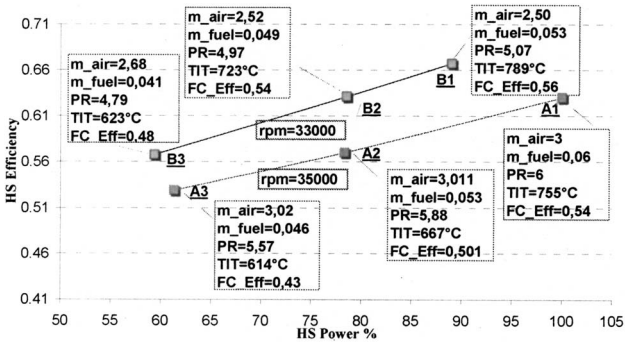


Fig. 9 Hybrid system part load efficiency

of the stack that operates at the same temperature with a lower value of the current density in comparison with the design point conditions [10].

Therefore, it is worthy to note that the ambient conditions must be very carefully taken into account also during the design point conditions definition. Moreover, an apt control of the cell temperature versus ambient temperature must be included in the whole control system.

5.2 Part-Load Conditions. In this case two different gas turbine control systems are considered, fixed and variable speed, as already discussed by the authors in [10,13].

5.3 Fixed Gas Turbine Rotational Speed. The only way to operate the system at a variable power output (part load operation), if a variable turbine speed control system is not available, is to vary the overall fuel flow rate feeding the plant. In all the calculations, the cell fuel utilization factor is kept as constant, and thus different fuel flow rates correspond to proportionally different electrical currents supplied by the SOFC stack. On the other hand, the air flow rate is an independent parameter, which is evaluated, as already said, on the basis of the matching between the expander, the compressor, and the stack, thus the oxygen utilization factor is not constant through all the simulation (see curves A1–A3 of Fig. 9).

In this case the cell efficiency and the HS efficiency decrease at part load operations due to the continuous decrease of the operating temperatures of the systems. From point A1 to A3 the new matching points require an increase of the air flow rate, a decrease of pressure ratio (but with an increase of the surge margin). In the same way fuel flow rate decreases, and also TIT decreases (about 150°C for a 35% power reduction). The impact of the part load conditions on the FC efficiency is also quite evident.

5.4 Variable Turbine Rotational Speed. The typical operation mode of gas turbine plants usually does not involve the possibility of changing the rotational speed of the turbine. The reason being that typical GT plants do not include an inverter, and thus the rotational speed of the turbine is chosen on the basis of the alternate current frequency required by the end user/electrical network. On the contrary, as the SOFC group of HS requires in any case the presence of an inverter (the inverter converts the electrical current produced by both the fuel cell and the alternator, the latter one being previously rectified) this configuration allows the operation of the turbine at variable rotational speed (variable frequency), taking also into account that for the size of the gas turbine used in this analysis the variable speed system may be considered as an available technology.

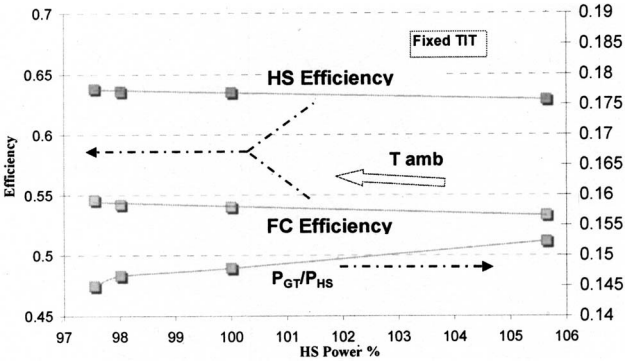


Fig. 8 Influence of ambient temperature with control system

In Fig. 9 the part-load performance of the IP-SOFC hybrid plant at different turbine rotational speeds is presented (curve B1–B3). The results show that the possibility of varying the rotational speed of the turbine is of fundamental importance in order to operate the plant at very high efficiency even at very low part load conditions (50% of the design point power). For each turbine rotational speed, the overall plant has been simulated between a maximum load obtained when the cell temperature is near to its allowed maximum value (higher load conditions have not been taken into account in order to respect this operating threshold). At the other extreme, the simulation was stopped when the temperature of the cell (and thus the turbine inlet temperature) are not compatible with the system feasible operating conditions.

Another important parameter that must be carefully checked at part load conditions is the compressor surge margin. While at fixed turbine speed this parameter increases reducing the HS load, using variable speed surge margin may decrease from one speed line to the next one, if the speed reduces, in particular if TIT (i.e., cell temperature) is maintained as a possible constant. Therefore, the compressor surge margin may be the limit for part load operation of the HS at variable speed, and the compressor must be carefully designed to reduce this impact.

Just to give an idea of the detail of the performance of the different parts of the plant at part-load conditions, here the attention has been restricted to the IP-SOFC group itself. The dashed curves drawn in Fig. 10 provide an explanation for the increase of cell efficiency by decreasing the power. In fact, along each dashed curve the cell operating temperature and the oxygen utilisation remain fairly constant whereas the current density decreases. As reported in a previous paper this new operating condition leads to a higher efficiency of the stack [10] and the operation at constant cell temperature can reduce the thermal stresses of the cell ceramic materials.

6 Conclusions and Future Work

In this paper the work carried out at Thermochemical Power Group (TPG) of University of Genoa in collaboration with Rolls Royce and Turbomeca for the development of a reliable design and off-design model of an IP-SOFC Hybrid System have been

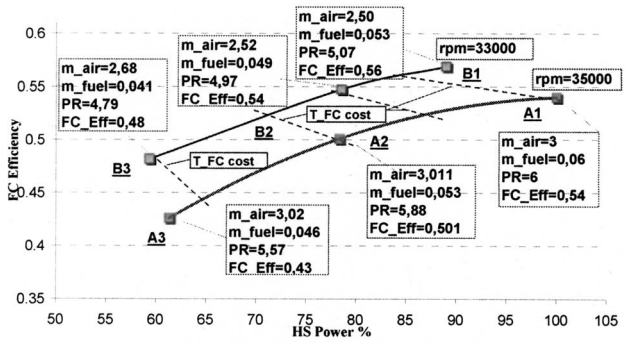


Fig. 10 Fuel cell part load efficiency

presented. The collaboration between the stack developer and TPG allows all the necessary data—geometrical, electrochemical, and experimental—to be easily available for the model development including their assessment. Moreover, turbomachinery maps of an about 400 kW gas turbine, made available by manufacturer, allowed an accurate representation of the machine off-design behavior in the HS operating conditions, improving overall system knowledge.

The work have been carried out as follows: (i) HS lay-out and design point definition using a generic IP-SOFC model; (ii) development and assessment of a detailed IP-SOFC model; (iii) refinement of the design point conditions using the HS model with detailed IP-SOFC one; (iv) ambient conditions influence investigation on IP-SOFC-HS performance; (v) part load performance of the IP-SOFC-HS at fixed and variable turbine speed. All the investigations have been carried out taking into account proper constraints for the stack, the turbine expander, and the compressor surge margin.

The main results of this work are:

1. The generic stack model is a very useful tool for preliminary investigations, particularly for the definition of the system layout and the preliminary optimization of the system parameters;
2. The detailed model, assessed with Rolls Royce IP-SOFC experimental data, has been included in the HS whole model to allow a better knowledge of the internal behavior of the cell both at design and off design conditions;
3. The design point conditions defined with the simplified IP-SOFC model have been modified to take into account the cell internal temperature distribution calculated with the detailed approach. In this way it is necessary to increase the air mass flow rate to obtain a feasible solution and as a result the system efficiency decreases;
4. Ambient conditions influence has been investigated in two different ways: (i) without modification and (ii) with modifications of cell parameters. In the first case the ambient temperature increase corresponds to an increase of the HS performance (just the opposite usually found for gas turbine systems). However, in this case the cell internal temperature distribution must be accurately verified. (ii) In the second case some modifications of the fuel flow rate and current density have been included in the calculation to avoid the cited unfeasible operative conditions. In this case the HS efficiency is about constant over the whole range of ambient temperature (the reduction of gas turbine power is compensated by the increase of cell efficiency and power). For the compressor surge margin no evident modifications are due to the ambient condition effects.
5. Part load conditions have been investigated using two different turbine control systems: (i) fixed speed, (ii) variable speed as already done by the author for Hybrid systems based on different cell technology [10].

The IP-SOFC-HS tools developed by TPG in collaboration with Rolls Royce are under continues development to include more details and find the optimum performance conditions of the IP-SOFC-HS under feasible operative range. In this way on going activities are the simulation of different size IP-SOFC-HS, start up and shut down analysis, transient and control system definition, etc.

Acknowledgment

This work has been developed within the Framework V-EESD project IM-SOFC-GT (EC contract ENK5-CT-2000-00302). The authors would like to thank all the partners of the project, and in particular Paola Costamagna of TPG, and Patrick Marconi of Turbomeca.

References

[1] Massardo, A. F., and Lubelli, F., 2000, "Internal Reforming Solid Oxide Fuel Cell-Gas Turbine Combined Cycles. Part A," ASME J. Eng. Gas Turbines Power, **122**, pp. 1–27.
[2] Campanari, S., and Macchi, E., 1999, ASME Paper No. 99-GT-84.
[3] Leo, A. J., Ghezel-Ayagh, H., and Sanderson, R., 2000, "Ultra High Efficiency Hybrid Direct Fuel Cell/Turbine Power Plant," ASME Paper No. 2000-GT-0552.
[4] Harvey, S. P., and Ritcher, H. J., 1993, "A Detailed Study of a Gas Turbine Cycle With an Integrated Internal Reforming Solid Oxide Fuel Cell," ASME J. Energy Resour. Technol.
[5] Layne, A., Samuelsen, S., Williams, M., and Holcombe, N., 2001, "Hybrid Fuel Cell Heat Engines: Recent Efforts," ASME Paper No. 2001-GT-0588.
[6] Bohn, D., Poppe, N., and Lepers, J., 2002, "Assessment of the Potential of Combined Micro Gas Turbine and High Temperature Fuel Cell Systems," ASME Paper No. GT-2002-30112.
[7] Costamagna, P., Bozzo, R., Magistri, L., and Massardo, A. F., 2002, "Simplified Versus Detailed Sofc Reactor Models and Influence on the Simulation of the Design Point Performance of Hybrid Systems," ASME Paper No. GT-2002-30653.
[8] Costamagna, P., Selimovic A., Bozzo, R., and Agnew, G., 2002, "Theoretical and Experimental Evaluation of a New Fuel Cell Concept: The Integrated Planar Solid Oxide Fuel Cell (IP-SOFC)," accepted for presentation at "ISCRE 17" Hong Kong, August 2002.
[9] Gardner, F. J., Day, M. J., Brandon, N. P., Pashley, M. N., and Cassidy, M., 1999, "SOFC Technology Development at Rolls-Royce," J. Power Sources, **86**.
[10] Costamagna, P., Magistri, L., and Massardo, A. F., 2001, "Design and Part-Load Performance of a Hybrid System Based on a Solid Oxide Fuel Cell Reactor and a Micro Gas Turbine," J. Power Sources, **96**, p. 352.
[11] Tarnowski, O. C., Agnew, G. D., Bozzolo, M., Massardo, A., Magistri, L., and Palsson, J., 2002, "Atmospheric and Pressurized Cycles for 1–2 MWe SOFC/GT Hybrid Systems," 5th European SOFC Forum, Luzern (CH), 1–5 July, 2002.
[12] Agnew, G., PWP, 2002, Fuel Cells for Marine Applications, ASME TURBOEXPO 2002, Amsterdam (NL).
[13] Agnew, G., Moritz, 2000, PowerGen Conference, USA.

Creep Strength and Microstructure of AL20-25+Nb Alloy Sheets and Foils for Advanced Microturbine Recuperators

P. J. Maziasz
e-mail: maziaszpj@ornl.gov

J. P. Shingledecker

N. D. Evans

Y. Yamamoto

K. L. More

R. Trejo

E. Lara-Curzio

Oak Ridge National Laboratory,
P.O. Box 2008,
Oak Ridge, TN 37831-6115

The Oak Ridge National Laboratory (ORNL) and ATI Allegheny Ludlum worked together on a collaborative program for about two years to produce a wide range of commercial sheets and foils of the new AL20-25+Nb™ (AL20–25+Nb) stainless alloy for advanced microturbine recuperator applications. There is a need for cost-effective sheets/foils with more performance and reliability at 650–750°C than 347 stainless steel, particularly for larger 200–250 kW microturbines. Phase 1 of this collaborative program produced the sheets and foils needed for manufacturing brazed plated-fin air cells, while Phase 2 provided foils for primary surface air cells, and did experiments on modified processing designed to change the microstructure of sheets and foils for improved creep-resistance. Phase 1 sheets and foils of AL20-25+Nb have much more creep-resistance than 347 steel at 700–750°C, and those foils are slightly stronger than HR120 and HR230. Results for Phase 2 showed nearly double the creep-rupture life of sheets at 750°C/100 MPa, and similar improvements in foils. Creep data show that Phase 2 foils of AL20-25+Nb alloy have creep resistance approaching that of alloy 625 foils. Testing at about 750°C in flowing turbine exhaust gas for 500 h in the ORNL Recuperator Test Facility shows that foils of AL20–25+Nb alloy have oxidation-resistance similar to HR120 alloy, and much better than 347 steel. [DOI: 10.1115/1.2718569]

1 Introduction

Previous reviews and papers adequately define (a) the critical role that recuperators play in advanced microturbines, (b) the growing evidence that 347 steel is inadequate above about 600°C, and (c) the need for the most cost-effective alloys which are capable of reliable performance at 650–700°C and above [1–6]. Use of opportunity or alternate fuels (gaseous or liquid) instead of clean natural gas, and cyclic operation will both further challenge recuperator reliability and durability.

While many recuperators are still being made from 347 stainless steel, several microturbine/recuperator manufacturers have been considering and testing sheets and foils of alloys with upgraded performance capabilities for a few years. Several years ago, ATI Allegheny Ludlum introduced the new AL20-25+Nb™ (AL20-25+Nb) stainless alloy as one such upgrade option for the primary-surface (PS) recuperator for the new Mercury 50 low-emission industrial gas turbine engine [7–9]. Oxidation testing in 7–10% water vapor at 700–760°C by Rakowski et al. [7–9] has shown good, long-term resistance of AL20-25+Nb stainless alloy to the moisture enhanced oxidation that plagues 347 stainless steels above 650°C. Similarly, good long-term resistance to moisture enhanced oxidation has been shown by Pint et al. [4,10] in 10% water vapor at 650–800°C for foils of HR120 and NF709 (also Fe-20Cr-25Ni alloy) alloys, far better than 347 steel. All investigators show outstanding resistance of the Ni-based superalloy 625 to moisture-enhanced oxidation. Currently, alloy 625 has been chosen as the upgrade alloy for the PS recuperator of the Solar Turbine Mercury 50 (4.6 MW) industrial gas turbine [5]. Kesseli et al. [3] have done cost/performance trade-offs in design studies of brazed-plate/fin (BPF) recuperator air cells for the Ingersoll-Rand PowerWorks 70 kW microturbine. Under the DOE Advanced Microturbine Materials Program, Lara-Curzio et al. [6] clearly demonstrate the significant shorter-term advantages of foils of HR120 alloy compared to 347 steel for both moisture-enhanced oxidation and creep in accelerated testing in the ORNL Recuperator Test Facility. Such in-turbine testing is particularly significant, because it also includes the dynamic effects of high-velocity gas during oxidation.

In 2004, ORNL and ATI Allegheny Ludlum began a collaborative project to extend previous work on the AL20-25+Nb stainless alloy. Phase 1 of this project produced a wide range of commercial quantities of the specific foils and sheets needed to manufacture air cells for both brazed plate and fin (BPF) and PS recuperators for the new, larger 200–250 kW advanced microturbine engines. Phase 2 of this project focused on modified processing of a limited quantity of sheets and foils to further improve creep-resistance. AL20-25+Nb alloy is another in the well-known group of austenitic stainless alloys based on the Fe-20Cr-25Ni composition, including NF709 from Nippon Steel Corp., and 12R72HV from Sandvik, which were developed to have more creep strength than 347 or 17-14CuMo stainless steels for fossil boiler tubing applications [11–14]. While alloy 625 has outstanding oxidation resistance and strength for advanced recuperator applications, its initial cost is still relatively high compared to 347 steel [3,15]. Recent price increases and surcharges for molybdenum and nickel have pushed the cost of alloys like 625 even higher. Alloy 625 also tends to have lower rupture ductility after aging or creep, particularly close to 700°C, which may also be a concern for cyclic microturbine operation. HR120 alloy has outstanding oxidation resistance, but it is not as strong as alloy 625

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received October 6, 2006; final manuscript received October 20, 2006. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2006: Land, Sea and Air (GT2006), May 8, 2006–May 11, 2006, Barcelona Spain. Paper No. GT2006-90195.

Table 1 Alloy compositions (wt %)

Alloy	Cr	Ni	Nb	N	C	Mo	Mn	Other
347	18	9.4	0.63	0.036	0.04	0.25	1.5	Bal. Fe
AL20-25 +Nb	20	26	0.38	0.1	0.09	1.5	1	Bal. Fe
HR120	25	35	0.7	0.2	0.05	2	1	Bal. Fe
625	21	bal.	3.4	···	0.025	8.3	0.03	4.4 Fe, 0.2 Ti, 0.2 Al
HR230	22	bal.	···	···	0.1	2	···	3 Fe, 14 W, 5 Co, 0.3 Al
PM2000	19	···	···	···	···	···	···	5.5 Al, 0.5 Ti, 0.5 Y_2O_3, bal. Fe

and costs only slightly less [15]. AL20-25+Nb alloy has the potential to cost significantly less than alloy 625, and thus to have the most cost-effective combination of oxidation and creep-resistance improvements relative to 347 steel. Specific changes to the foil/sheet processing parameters produced the Phase 2 AL20-25+Nb alloy, with small, controlled increases in the average grain size, as well as refinements in the nanodispersions of NbC precipitation for better creep-resistance. The technical basis for such foil/sheet processing changes was established by previous ORNL/ATI Allegheny Ludlum work that produced AL347HP™ [15,16].

2 Experimental Details

This collaborative project began with a larger amount of hot-band from the original heat of AL20-25+Nb, and the alloy composition is given in Table 1. Compositions of heats of 347 steel and alloy 625 supplied as foils and sheets by ATI Allegheny Ludlum to compliment this study are also listed in Table 1. Compositions of all other alloys tested at ORNL and used to make further comparisons are similar to or the same as those reported previously [15].

This collaborative project was divided into two phases. Phase 1 was the main commercial foil and sheet effort, producing over 1000 lb (453.6 kg) of finished foils, and 800 lb (363 kg) of finished sheets, with standard processing similar to previous work on the AL20-25+Nb alloy [7,9]. Phase 2 of this project produced a much smaller quantity, and only selected thicknesses of foils or sheets, but involved important changes in the processing parameters specifically designed to modify the grain size and solutionize NbC for improved creep resistance.

Phase 1 of this project at ATI Allegheny Ludlum processed about 5000 lb (2,268 kg) of original hot-band into over 1000 lb (453.6 kg) of foils (0.004, 0.005, and 0.008 in. thickness) (0.1, 0.127, and 0.2 mm, respectively), and 800 lb (363 kg) of sheets (0.010 and 0.015 in. thicknesses) (0.41 and 0.81 mm) with standard processing conditions. Portions of the 0.010 (0.41 mm) and 0.015 in. (0.81 mm) thick sheets were slit into final finished coils, while other portions were used as intermediate stock for further reductions to foils. The Phase I foil/sheet products were slit and shipped to Ingersoll Rand Energy Systems (IRES) in 2004/2005, to produce the various different pieces required for making brazed plate-and-fin (BPF) air cells for recuperators for their new 250 kW PowerWorks microturbine.

Phase 2 of this effort modified the processing of 0.010 in. (0.41 mm) sheet, and a portion of that was further reduced (with modified processing at this step too) to make several hundred pounds of 0.0032 in. (0.08 mm) finished foil, which was slit and shipped to Capstone Turbines (CT) for manufacturing trials for welded primary-surface (PS) air cells for recuperator application in support of their 200 kW microturbine. The remainder of the Phase 2 effort retained some of the 0.010 in. (0.41 mm) thick sheet for creep properties studies, and reduced a portion into 0.005 in. (0.127 mm) thick foil, with further processing modification for enhanced creep resistance.

Panels of all the various Phase 1 and Phase 2 sheets and foils of AL20−25+Nb were shipped to ORNL for more detailed creep-rupture and lab-scale oxidation testing, as well as detailed microstructural analysis. Some of these foils have been used to make probes for testing in the ORNL Recuperator Testing Facility, and some initial test results are included here, while longer-term testing is in progress. ORNL wire-EDMed out special sheet creep specimens from each foil or sheet panel, with the specimen gauge length parallel to the rolling direction of the material. Tabs were also cut from sheets and foils, and tack-welded to the shoulders for support in the grips. Creep-rupture testing was performed in dead-load machines, with furnaces opened to air. Extensometers are used to measure strain, with computer controlled digital data acquisition. Generally only one specimen per alloy/condition was creep-tested due to limitations in available equipment. However, in some cases, duplicate creep tests were run.

Microstructural analysis included standard optical metallographic techniques to reveal grain size for transverse cross sections parallel to the rolling direction of all the as-processed foils and sheets. Some of the creep-ruptured specimens were also examined in transverse cross section by optical metallography and scanning electron microscopy (SEM). High resolution SEM was performed on some of the creep-ruptured metallographic specimens (polished and unetched) or on electropolished transmission electron microscopy (TEM) discs EDMed from the gauge portion of the creep specimen, away from the fracture area. Phase identification was done on a JEOL 6500F microscope, with a field emission gun (FEG) and capability for both secondary electron and backscattered electron imaging (SE and BSE, respectively). BSE imaging was done at 5 kV for enhanced contrast, while x-ray energy dispersive spectroscopy (XEDS) was done at 15 kV for microcompositional analysis of individual precipitate particles.

Results and Discussion

Creep-Rupture Testing. Figures 1 and 2 contain updates of the ongoing long-term creep-rupture testing of results presented last year [5], at 750°C/100 MPa and 704°C/152 MPa, respectively. Alloy 625 commercial sheet (0.015 in. or 0.41 mm thick) shows excellent creep-resistance at both 704 and 750°C, with only about 2% strain and no rupture after over 13,000 h at 750°C. Sheet of the Fe-based oxide dispersion-strengthened (ODS) Plansee alloy, PM2000, by contrast shows minimal creep (<0.2%) after about 12,000 h at a higher stress; however, PM2000 is much more expensive and more difficult to fabricate for these recuperator applications. Figures 1 and 2 show new data for alloy 625 foils (0.004 in. or 0.1 mm thick) at both 704 and 750°C, and clearly show that such foils have much less creep-resistance than sheets, suggesting an effect of fine grain size reducing creep-rupture resistance, as is found in many other heat-resistant alloys [15]. Measurements show that the grain size of the alloy 625 foil is much finer than sheet (8.7 compared to 39 μm). However, at 750°C, the foil of alloy 625 is still much more creep-resistant than foil of another Ni-based superalloy, HR230. Figure 2 also shows that at

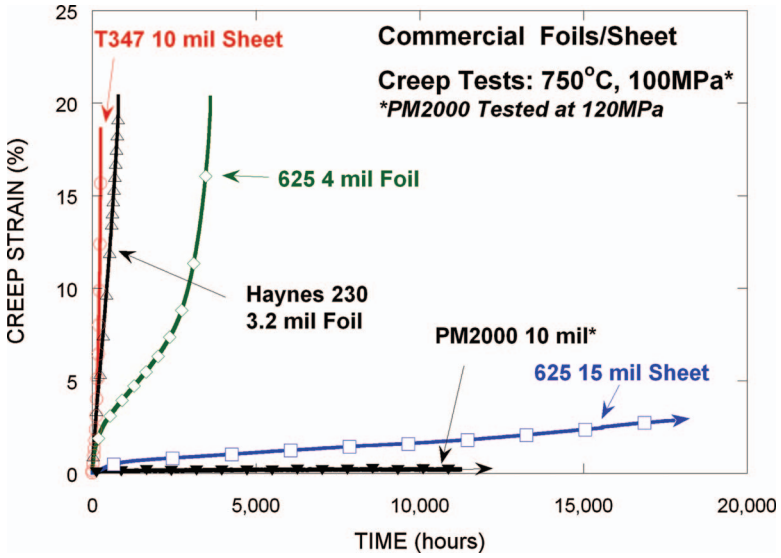


Fig. 1 Creep-rupture strain versus time data for various commercial heat-resistant alloy sheets and foils tested at 750°C and 100 MPa (except for the PM2000 ODS alloy, tested at 120 MPa) in air. Both alloy 625 and the PM2000 (ODS ferritic alloy) are still in-test.

704°C, AL20-25+Nb (Phase 1) alloy foil has similar creep-resistance but over twice the rupture ductility compared to the HR120 (Fe-25Cr-35Ni, Nb, Mo, C, N) stainless alloy.

Results of all ORNL creep-rupture testing done at 650–750°C to date on foils (0.003–0.005 in. thick) of AL20-25+Nb alloy and other heat-resistant alloys considered or used for recuperator applications are compared in Fig. 3, using the Larson-Miller parameter, calculated using rupture time and test temperature. For foils, all of these more heat-resistant alloys have significantly more creep-resistance than standard 347 steel. Unfortunately, processing to control grain size and intragranular NbC dispersions in 347 steel for improved creep-rupture resistance [15] does nothing to change its inadequate resistance to moisture-enhanced oxidation [10], which is a function of the alloy composition, particularly Cr and Ni contents. AL20-25+Nb, HR120, HR230, and 625 alloys are all inherently more resistant to moisture-enhance oxidation due to both higher Cr and Ni alloying element contents. The creep-rupture strength of AL20-25+Nb (Phase 1) foils is slightly higher than HR230 and HR120 alloy foils. Since HR230 is almost twice the cost of HR120, both HR120 and AL20-25+Nb are more cost-effective for the same level of creep-resistance. The initial testing on foils of AL20-25+Nb (Phase 2) alloys is not complete, and those results show significantly more creep resistance than comparable Phase 1 material; the creep-resistance of AL20-25+Nb alloy (Phase 2) foils now comes closer to the behavior of alloy 625 than any of the other alloys.

Figure 4 shows a more detailed comparison of individual creep strain curves at 750°C/100 MPa for AL20-25+Nb alloy sheets with Phase 1 and Phase 2 processing, while Fig. 5 shows a similar comparison for foils. Sheets of 347 steel typically have more creep resistance than thin foils, with 200–250 h compared to 50 h or less for standard processing [15]. AL20-25+Nb (Phase 1) sheet has similar creep resistance to 347 steel (similar creep rate), but has a longer rupture life by virtue of having more creep ductility before rupture (Fig. 4). By comparison, AL20-25+Nb (Phase 2)

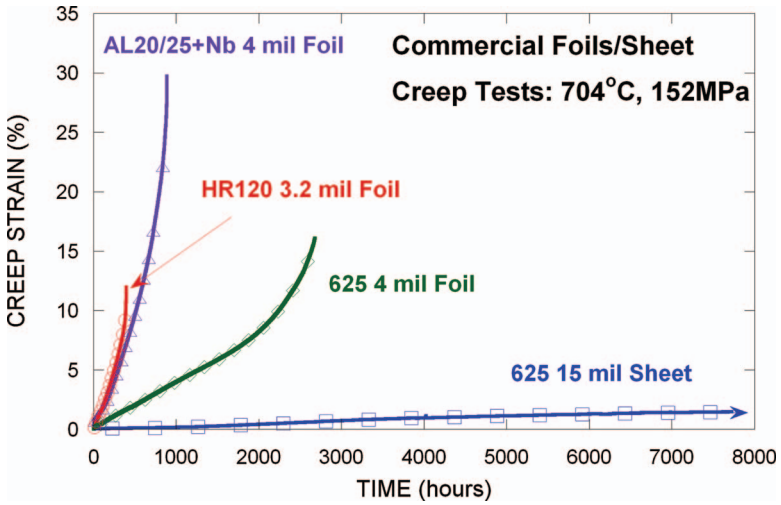


Fig. 2 Creep data on commercial sheets and foils tested at 704°C and 152 MPa in air. The sheet of alloy 625 has shown little creep at these conditions so far, and is still in-test.

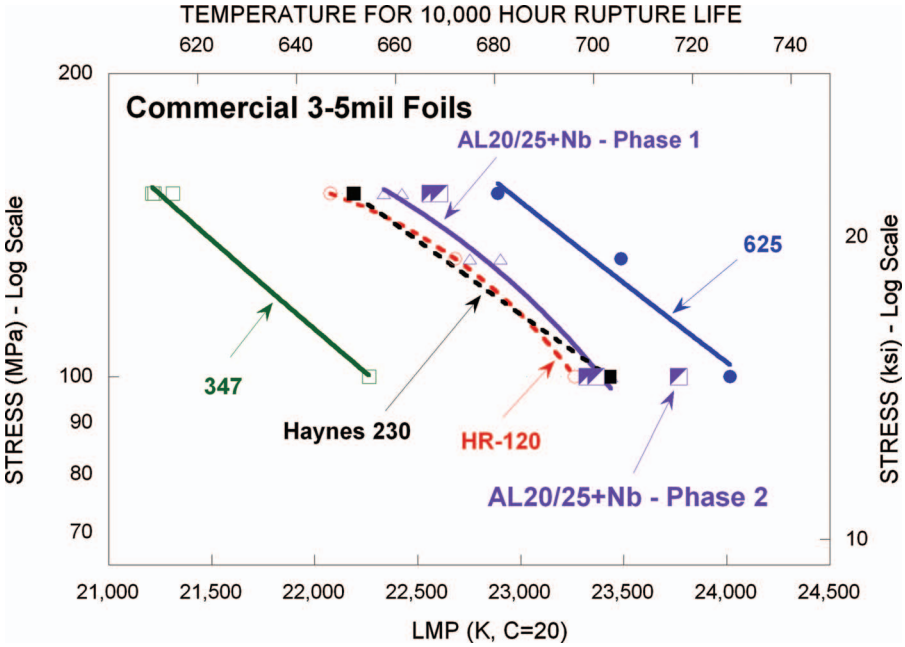


Fig. 3 A plot of creep-rupture stress versus Larson-Miller parameter (LMP) for the various foils of commercial heat-resistant stainless steels, stainless alloys, and Ni-based superalloys tested at 650–750°C in air at ORNL

alloy sheet has about double the rupture life due to a much lower creep rate and slightly better rupture ductility. Clearly, even a small change in processing has a significant effect on creep-resistance for this alloy. Figure 5 indicates a far more dramatic effect of processing changes on AL20-25+Nb, despite slight differences in foil thickness. With Phase 1 processing, 0.004–0.005 in. (0.1–0.127 mm) thick foils of AL20-25+Nb alloy have more creep resistance than 347 steel, but have slightly higher creep rates than HR120, and slightly longer rupture lives due to almost three times the rupture ductility. The 0.0032 in. (0.08 mm) thick foil of AL20-25+Nb (Phase 2) alloy shows over three times the creep-rupture life of HR120 alloy, due to the dramatically prolonged secondary creep regime at the minimum creep rate combined with double the rupture ductility. The effect of changing the processing parameters has an even more dramatic

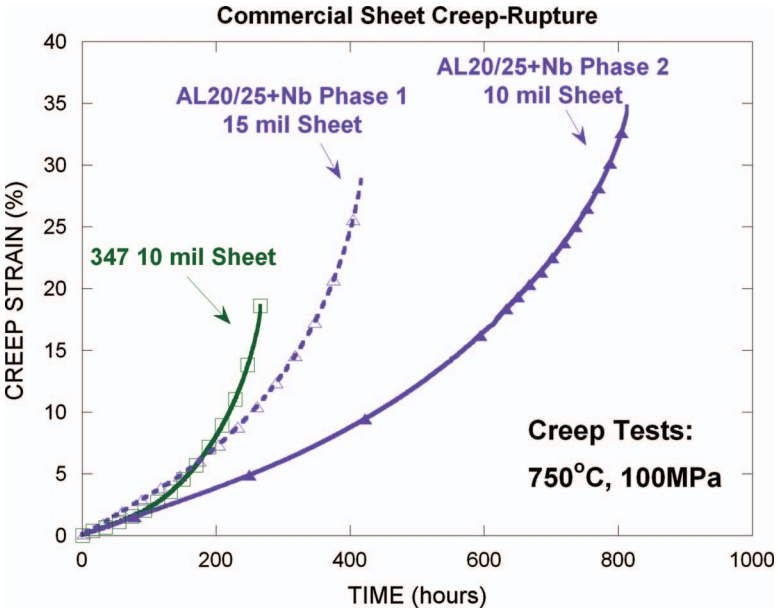


Fig. 4 Comparison of creep-strain versus time plots for creep-rupture testing of 10–15 mil sheets of standard 347 stainless steel, and of AL20-25+Nb stainless alloy at 750°C and 100 MPa in air. Differences in processing parameters produced coarser uniform grain size in the Phase 2 material compared to the Phase 1 material.

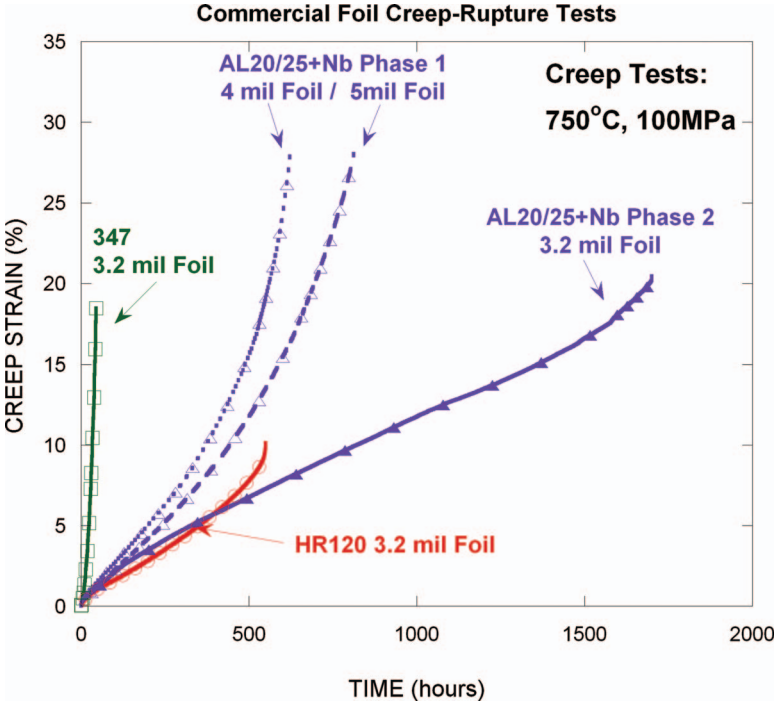


Fig. 5 Comparison of creep-strain versus time plots for creep-rupture testing of 3–5 mil foils of standard 347 stainless steel, and of HR120 and AL20-25+Nb stainless alloys at 750°C and 100 MPa in air. Differences in processing parameters of the AL20-25+Nb alloy produced a coarser uniform grain size in the Phase 2 material compared to the Phase 1 material.

effect on foils of AL20-25+Nb alloy than it does on thicker sheets. Similar creep-rupture testing of 0.005 in. (0.127 mm) thick foil of Phase 2 material is nearly complete to verify this behavior.

Microstructural Analysis of Creep-Tested Foils. Microstructural analyses of as-processed grain size structures of Phase 1 and Phase 2 sheets (0.010 in. or 0.41 mm thick) of AL20-25+Nb are shown in Fig. 6. Both the metallographic images and the grain size distribution histographs show that Phase 2 processing increases the uniform grain size by coalescence of most of the patches of finer grains that are present in the Phase 1 microstructure. Average grain size numbers are presented in Table 2 for all the AL20-25+Nb alloy foils and sheets for both Phase 1 and Phase 2 materials. The change in average grain size is more dramatic for the 0.010 in. (0.41 mm) thick sheet specimens than the 0.005 in. (0.127 mm) thick foils, but that is likely due to the sheets containing larger regions of very fine grains (grains $<10-12$ μm), which are removed by the higher temperature processing. For reference, a grain size of about 12 μm corresponds to an ASTM grain size of 9.5, whereas a grain size of 4 μm corresponds to an ASMT grain size of 12.5. By comparison to several of the other heat-resistant alloy foils, the 625 foil has a grain size similar to AL20-25+Nb alloy, whereas the HR120 foil has a much coarser grain size, close to 28 μm (ASTM grain size of about 7).

Comparison of the microstructures of gauge cross sections from AL20-25+Nb alloy sheets of Phase 1 and Phase 2 materials after creep-rupture at 750°C/100 MPa are shown in Fig. 7 (these are taken from the creep-tested specimens shown in Fig. 4). The cavities or porosity seen in both samples are the creep damage caused by creep-voids forming at grain boundaries and triple points, which can then link up to separate the grain boundaries and weaken the material and cause final fracture. Clearly the Phase 1 materials have many more such internal cracks, with thinning of the areas with more cracks because they are now weaker. By contrast, the Phase 2 material, which is thinner sheet and has also been exposed to those same creep conditions and lasted twice as long, has less of such creep damage, and that damage is more confined to the center of the specimen. While TEM analysis has not been done yet, this does suggest that the processing differences have at least retarded the mechanisms causing creep damage, and possibly have changed them. More detailed microstructural analysis is needed.

Finally, higher magnification analysis of grain boundaries and precipitation in the gauge sections of AL20-25+Nb (Phase 2) and HR120 foils tested at 750°C/100 MPa are shown in Fig. 8. While both alloys show mixtures of Si-Mo-Cr-Ni M_6C and Cr-rich $M_{23}C_6$ particles along the grain boundaries and in the matrix, there are some obvious differences. The HR120 has a finer dispersion of mainly $M_{23}C_6$ and some M_6C particles along the grain boundaries, while the AL20-25+Nb has coarser $M_{23}C_6$ and M_6C, with a larger relative fraction of the latter. The AL20-25+Nb alloy also shows the unexpected presence of fine laths of AlN engulfed in many of the Cr-rich carbides. This most likely reflects Al contamination of this particular first large, developmental heat. This should not be typical of other commercial production heats of AL20-25+Nb alloy, where such Al contamination effects have been eliminated [16].

Initial Test Results From the ORNL Recuperator Tesing Facility. Sample holders for the ORNL Recuperator Testing Facility were made by welding foils (0.089 mm or 0.0035 in. thick) to the four different positions, and then pressurizing them with air inside. Thermocouples were attached to measure temperature. More details on the development of this facility and the test results for a wider range of alloys are available elsewhere [6,17]. Holders made with standard 347 steel, with HR120 alloy, and with AL20-25+Nb alloy were inserted into the modified recuperator of a Capstone 60 kW microturbine, and then exposed to flowing exhaust gas for 500 h at temperatures of 630–750°C. Scanning electron microscope (SEM) analysis of foil cross-section specimens from the hottest position (about 750°C) are shown in Figs.

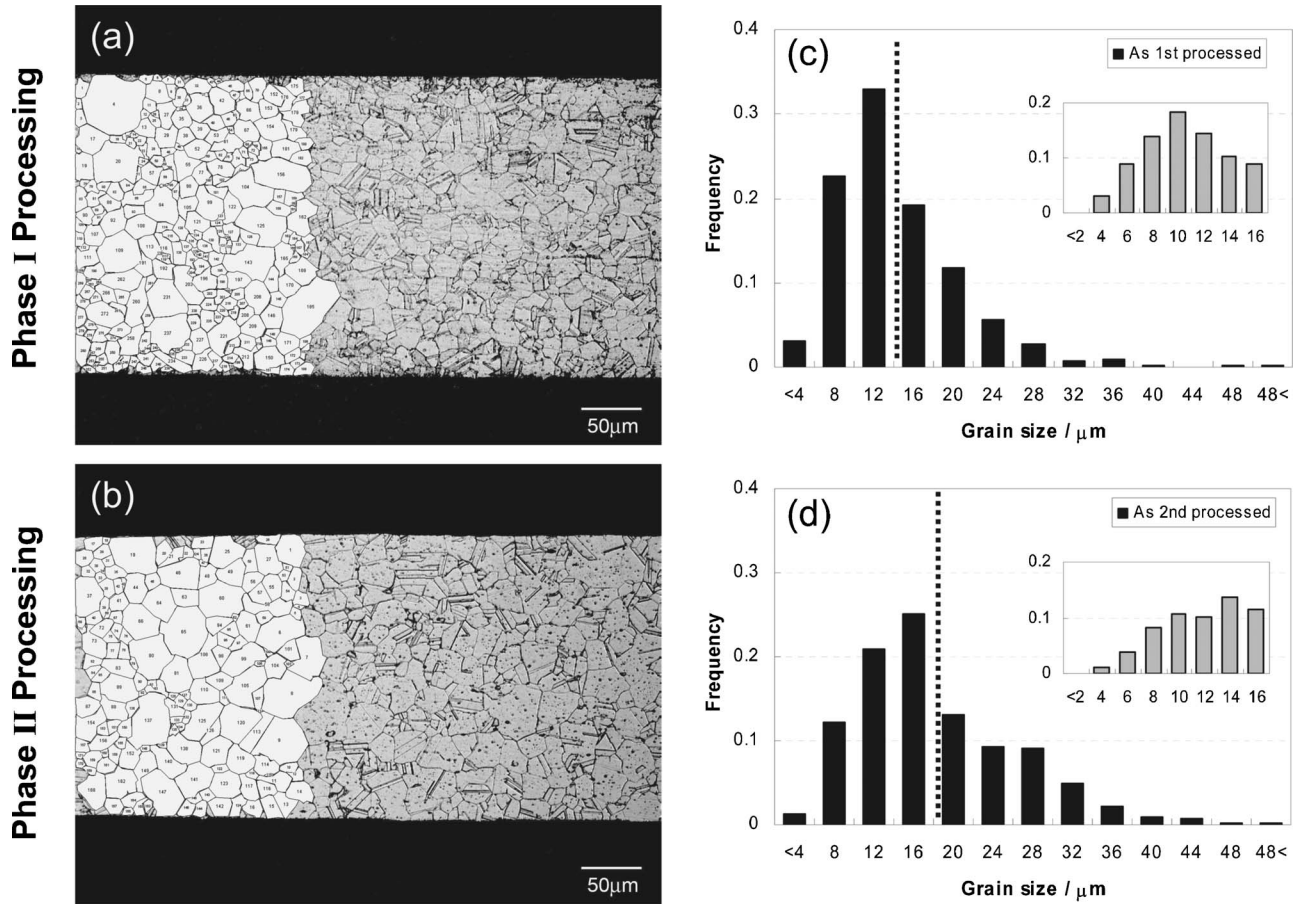


Fig. 6 Optical metallographic microstructures of cross-section specimens of AL20-25+Nb sheets (0.010 in. or 0.254 mm thick) from (*a*) Phase 1 processing, and (*b*) Phase 2 processing. Superimposed on the left hand portions of (*a*) and (*b*) are redrawn schematic diagrams of the actual grain structure that were used for quantitative grain size distribution analysis. Histographs of the grain size distribution obtained from analysis of those images are plotted in (*c*) Phase 1 and (*d*) Phase 2. Dotted lines on these histrographs indicate the average grain size, and the smaller inset histographs are analyses of the smallest grains using finer grain-size steps.

9(*a*)–9(*c*) for each of these alloys, while the corresponding macroscopic exposed foil surfaces are shown in Figs. 9(*d*)–9(*f*). The 347 steel showed severe moisture-enhanced breakaway oxidation after only 500 h, which was clearly worse on the turbine exhaust-gas side relative to the air side (Figs. 9(*a*) and 9(*d*)). The outer 66% of the 20 μm thick oxide on the 347 steel is mainly Fe-rich oxide, with a buried Cr-rich oxide layer between the oxide scale and the metal. Oxidation has consumed about 25% of the original foil thickness of the 347 steel after only 500 h. The 347 steel foil also macroscopically exhibits bulging (indicative of creep) and oxide spallation (Fig. 9(*d*)). By contrast, both the HR120 and AL20-25+Nb alloys show good oxidation-resistance at this point, with thin, protective Cr-rich oxide scales (Figs. 9(*b*) and 9(*c*)), and

Table 2 Quantitative grain size data for AL20-25+Nb alloy

	Average grain size/μm	
Thickness	Phase 1	Phase 2
15 mil (381 μm)	11.4	···
10 mil (254 μm)	12.8	18.2
8 mil (203 μm)	11.4	···
5 mil (127 μm)	11.2	12.5
4 mil (102 μm)	12.5	···
3.2 mil (81.3 μm)	···	13.2

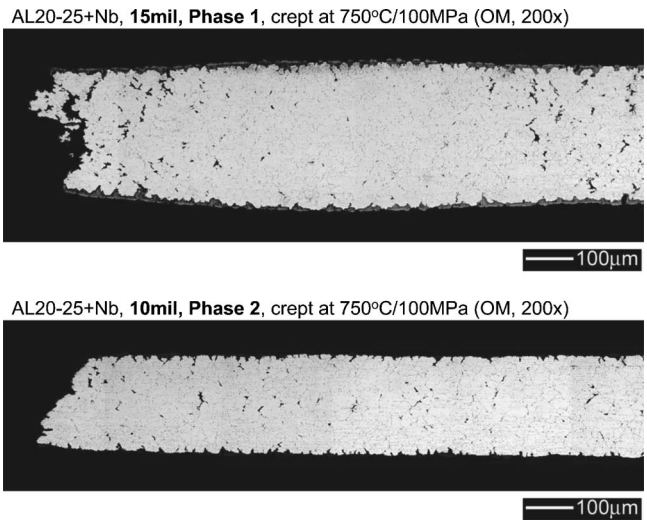


Fig. 7 Optical metallographic micrographs showing the transverse cross-sections sheet specimens of sheets of AL20-25+Nb after creep-rupture testing at 750°C and 100 MPa in air. The top sheet specimen is 15 mil Phase 1 alloy, which ruptured after about 400 h, while the bottom sheet is 10 mil Phase 2 alloy, which ruptured after over 800 h. The creep-rupture curves corresponding to these specimens are shown in Fig. 4.

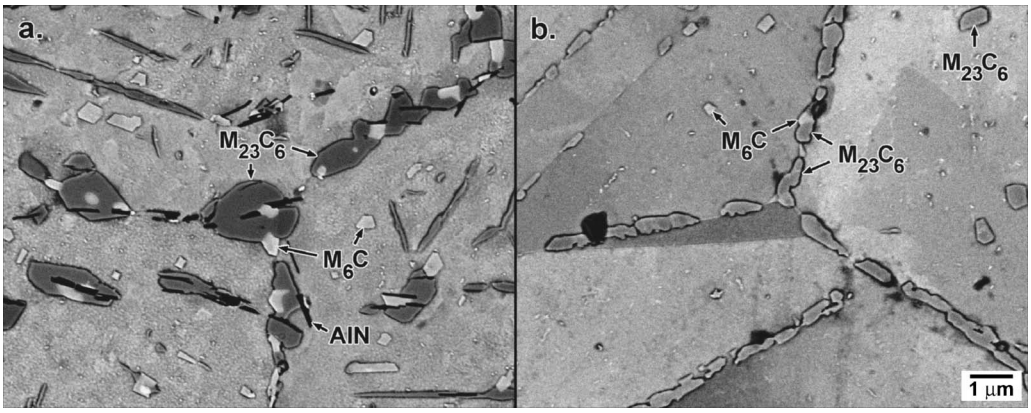


Fig. 8 Scanning electron microscopy back-scattered (SEM/BS) images of typical grain boundary regions showing the differences in precipitate phases developed during creep-rupture testing at 750°C and 100 MPa in 3.2 mil foils of (*a*) AL20-25+Nb stainless alloy, Phase 2 processing, with rupture after about 1700 h, and (*b*) HR120, with rupture after about 530 h. Both alloys contain mixtures of Cr-rich $M_{23}C_6$ and Mo-, Si-, and Ni-rich M_6C phase along the grain boundaries and within the grains, and both pictures are at the same magnification. Creep curves corresponding to both of these specimens are shown in Fig. 5.

macroscopically show no evidence of any break-away oxidation, and no evidence of bulging (Figs. 9(*e*) and 9(*f*)). Exposure of these foils at 750°C can be considered accelerated testing for longer-term exposures at lower temperatures. Both the HR120 and AL20-25+Nb stainless alloys show much better degredation (oxidation and creep-bulging) resistance compared to standard 347 stainless steel, and should have longer lifetimes and higher temperature capability for comparable recuperator applications. Matthews [18] shows dramatic improvement of PS air cells made of HR120 foils compared to 347 steel, when exposed in a Capstone

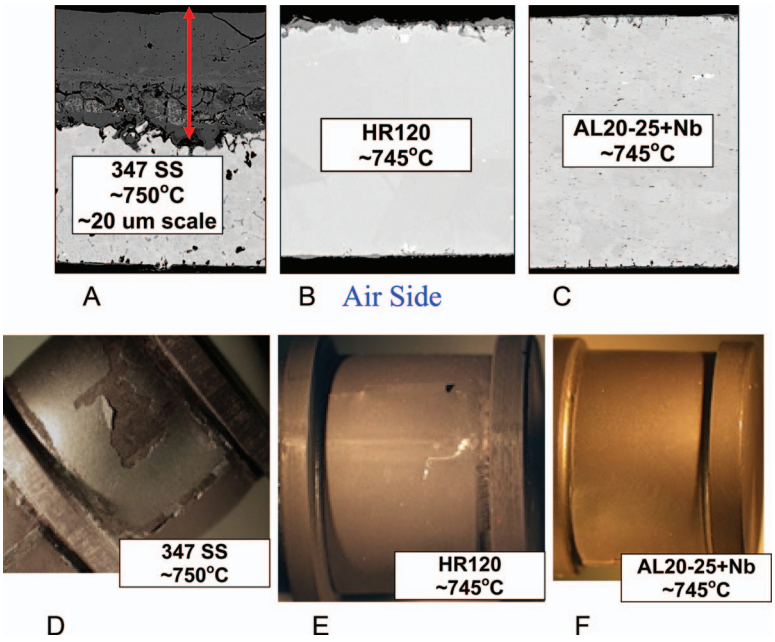


Fig. 9 Sample holders for the ORNL Recuperator Testing Facility were made by welding foils (0.089 mm or 0.0035 in. thick) to the various positions, and then pressurizing them with air inside. Holders made with standard 347 steel, with HR120 alloy, and with AL20-25+Nb alloy were inserted into the recuperator of a modified Capstone 60 kW microturbine, and exposed to flowing exhaust gas for 500 h. Scanning electron microscope (SEM) analysis of foil cross-section specimens from the hottest position are shown in (*A*)–(*C*) for each of the alloys, while the macroscopic exposed foil surfaces are shown in (*D*)–(*F*). The 347 steel experiences very severe moisture-enhanced oxidation attack in the exhaust gas at about 750°C (*A*), and microscopically exhibits bulging (indicative of creep) and oxide spallation (*D*). By contrast, both the HR120 and AL20-25+Nb alloys show good oxidation-resistance at this point, with thin, protective Cr-rich oxide scales (*B*) and (*C*), and clean looking surfaces with no evidence of bulging.

60 kW MicroTurbine run for 1800–2500 h at 100°F above the turbine inlet temperature (TET). Specimens of AL20-25+Nb are currently being exposed for 3000–5000 h tests in the ORNL Recuperator Testing Facility, and will be analyzed at a later date.

Preliminary Feedback on Recuperator Manufacturing. Ingersoll-Rand Energy Systems (IRES) has provided some preliminary, qualitative feedback on the use of the new AL20-25+Nb (Phase 2) alloy foils and sheets for manufacturing tests prior to making BPF air cells. While PS air cells tend to be made from the same foils of the same alloy, BPF air cells can introduce the more heat/corrosion resistant alloys for only those parts that see the highest temperatures, resulting in a mixed-alloy air cell. To date, IRES indicates the new alloy has good folding, forming, welding, and brazing characteristics, similar to 347 steel. Capstone Turbines has also received foil of AL20-25+Nb (Phase 2) alloy for testing and evaluation for their PS air-cell manufacturing requirements, which include folding and welding. More test results will be shared as they become available.

Summary

A collaborative project by ORNL and ATI Allegheny Ludlum in Phase 1 has produced a wide range of commercial foils and sheets of the new AL20-25+Nb alloy for recuperator air-cell manufacturing with upgraded performance relative to 347 stainless steel. These foils and sheets show better creep resistance than 347 steel at 700–750°C, and Phase 1 foils meet or exceed the creep-rupture strength of HR120 and HR230. Phase 2 processing increased the average grain size and eliminated the smallest grains to boost the creep-resistance significantly. Phase 2 foils of the AL20-25+Nb alloy come closer to the strength of alloy 625. Creep-testing of the Phase 1 materials has been completed, and similar testing of the Phase 2 material is also nearly finished. Initial testing of the AL20-25+Nb alloy for 500 h at up to 750°C in the ORNL Recuperator Testing Facility shows dramatically improved resistance to moisture-enhanced oxidation and improved creep-resistance relative to standard 347 stainless steel. The behavior of the AL20-25+Nb is similar to that shown for HR120 alloy, for similar in-turbine testing after 500 h.

Both HR120 and AL20-25+Nb alloys are expected to provide microturbine recuperators with all of the fabrication benefits of 347 steel (easy folding, welding, and brazing), and still provide performance and durability improvements, even with higher recuperator temperatures, cyclic operation or alternative, more corrosive opportunity fuels (flare gas, land-fill gas, biofuels, etc.).

Acknowledgments

Many thanks to Dr. Charles P. Stinner of the ATI Allegheny Ludlum Technical Center, who helped initiate this collaborative project, worked closely with others in production facilities, and who also helped in some of the technical analyses. Cooperation of Capstone Turbines, Inc., and Ingersoll-Rand Energy Systems, with the materials research efforts at ORNL is greatly appreciated. Research was sponsored by the U.S. Department of Energy, Assistant Secretary for Energy Efficiency and Renewable Energy, Office of Distributed Energy and Electrical Reliability, and at the SHaRE User Facility by the Division of Materials Sciences, Office of Basic Energy Sciences, under Contract No. DE-AC05-00R22725 with UT-Battelle, LLC.

References

[1] McDonald, C. F., 2003, "Recuperator Considerations for Future Higher Efficiency Microturbines," Appl. Therm. Eng., **23**, pp. 1463–1487.
[2] Treece, B., Vessa, P., and McKeirnan, R., 2002, "Microturbine Recuperator Manufacturing and Operating Experience," ASME Paper No. GT-2002-30404.
[3] Kesseli, J., Wolf, T., Nash, J., and Freedman, S., 2003, "Micro, Industrial, and Advanced Gas Turbines Employing Recuperators," ASME Paper No. GT2003-38938.
[4] Pint, B. A., and More, K. L., 2004, "Stainless Steels With Improved Oxidation Resistance for Recuperators," ASME Paper No. GT2004-53627.
[5] Maziasz, P. J., Shingledecker, J. P., Pint, B. A., Evans, N. D., Yamamoto, Y., More, K. L., and Lara-Curzio, E., 2005, "Overview of Creep Strength and Oxidation of Heat-Resistant Alloy Sheets and Foils for Compact Heat-Exchangers," ASME Paper No. GT2005-68927.
[6] Lara-Curzio, E., Trejo, R., More, K. L., Maziasz, P. J., and Pint, B. A., 2005, "Evaluation and Characterization of Iron- and Nickel-Based Alloys for Microturbine Recuperators," ASME Paper No. GT2005-68630.
[7] Rakowski, J. M., Stinner, C. P., Lipschutz, M., and Montague, J. P., 2004, "The Use and Performance of Oxidation and Creep-Resistant Stainless Steels in an Exhaust Gas Primary Surface Recuperator Application," ASME Paper No. GT2004-53917.
[8] Stambler, I., 2004, "Mercury 50 Rated at 4600 kW and 38.5% Efficiency With 5 ppm NOx," Gas Turbine World (Feb.-Mar.), pp. 12–16.
[9] Rakowski, J. M., Stinner, C. P., Bergstrom, D. S., Lipschutz, M., and Montague, J. P., 2005, "Performance of Oxidation and Creep Resistant Alloys for Primary Surface Recuperators for the Mercury 50 Gas Turbine," ASME Paper No. GT2005-68313.
[10] Pint, B. A., 2005, "The Effect of Water Vapor on Cr Depletion in Advanced Recuperator Alloys," ASME Paper No. GT2005-68495.
[11] Staubli, M., et al., "Materials for Advanced Steam Power Plants: The European COST522 Action," *Parsons 2003: Engineering Issues in Turbine Machinery, Power Plants and Renewables*, The Institute of Materials, Minerals and Mining, Maney, London, pp. 305–324.
[12] Kikuchi, M., Sakakibara, M., Otoguro, Y., Mimura, H., Araki, S., and Fujita, T., 1987, "An Austenitic Heat Resisting Steel Tube Developed For Advanced Fossil-Fired Steam Plants," *High Temperature Alloys, Their Exploitable Potential*, Elsevier Science, New York, pp. 267–276.
[13] Takahashi, T., et al., 1988, "Development of High-Strength 20Cr-25Ni (NF709) Steel for USC Boiler Tubes," Nippon Steel Technical Report No. 38, Nippon Steel Corp., Tokyo, Japan.
[14] *Quality and Properties of NF709 Austenitic Stainless Steel for Boiler Tubing Applications*, 1996, Nippon Steel Corp., Revision 1.1, Tokyo, Japan.
[15] Maziasz, P. J., Swindeman, R. W., Shingledecker, J. P., More, K. L., Pint, B. A., Lara-Curzio, E., and Evans, N. D., 2003, "Improving High Temperature Performance of Austenitic Stainless Steels for Advanced Microturbine Recuperators," *Parsons 2003: Engineering Issues in Turbine Machinery, Power Plants and Renewables*, The Institute of Materials, Minerals and Mining, Maney, London, pp. 1057–1073.
[16] Stinner, C., 2006, "Processing to Improve Creep and Stress Rupture Properties of Alloy T347 Foil," 2003, Allegheny Ludlum Technical Center Internal Report, Brackenridge, PA, available upon request; also private communication about differences in trace aluminum contamination in subsequent commercial heats of AL20-25+Nb alloy.
[17] Lara-Curzio, E., Trejo, R., More, K. L., Maziasz, P. J., and Pint, B. A., 2004, "Screening and Evaluation of Materials for Microturbine Recuperators," ASME Paper No. GT2004-54254.
[18] Matthews, W. J., 2006, "Additional Engine Testing of an Advanced Alloy for Microturbine Primary Surface Recuperators," ASME Paper No. GT2006-90068.

Control Options for Load Rejection in a Three-Shaft Closed Cycle Gas Turbine Power Plant

B. W. Botha
Senior Lecturer
e-mail: barend.botha@nwu.ac.za

P. G. Rousseau
Professor

School of Mechanical Engineering,
North-West University,
Potchefstroom Campus,
Private Bag X6001,
Potchefstroom 2520, South Africa

An important issue to be addressed in power plants is the continued operation during load transients, such as load following and load rejection. It is inevitable that with new power plant technology, new control strategies will be required. One such technology investigated for commercial power plants is that of a three-shaft recuperative intercooled closed-loop Brayton cycle with a high-temperature gas-cooled nuclear reactor as the heat source and helium as coolant. Because of its unique configuration, the utilization of traditional power plant control strategies is limited. In order to address this, detailed cycle analyses were performed to identify new potential control strategies. The analyses were done using the Flownex thermohydraulic systems CFD simulation software since it is ideally suited for component and system integration. It also enables designers to simulate complex load scenarios and design-suitable controller algorithms. It was therefore possible to investigate control options for one of the most severe load control scenarios, i.e., that of full load rejection due to the loss of the grid power. This paper briefly describes the various control strategies investigated and presents details of the two strategies showing the most promising results with regard to load rejection.
[DOI: 10.1115/1.2718225]

Introduction

The three-shaft, high-temperature gas-cooled reactor cycle layout used in this study was based on the cycle shown in Fig. 1. The cycle primarily consists of a precooler (PC), low-pressure compressor (LPC), intercooler (IC), high-pressure compressor (HPC), recuperator heat exchanger (RX), pebble bed nuclear reactor, high pressure turbine (HPT), low-pressure turbine (LPT), and power turbine (PT). Cooling is supplied via an intermediate cycle using desalinated water to exchange heat with an external heat sink, which is intended to be seawater.

Apart from the components mentioned above, Fig. 1 also shows all the components used for the various control options investigated during the study. These include variable turbine (HPTV, LPTV, and PTV) and compressor (HPCV and LPCV) guide vanes as well as a number of valves. The different valves used include the local compressor by-pass valves (HPCV and LPCV), the compressor by-pass valve (CPBP), recuperator by-pass valve (RXBP), power turbine by-pass (PTBP) valve, and power turbine in-line valve (PTIL). Finally, some options also have a resistor bank connected to the generator to dissipate excess power generated by the turbine.

An example of such a three-shaft system is the so-called pebble bed micromodel (PBMM) [1] that was designed and built by the Faculty of Engineering at the Potchefstroom Campus of the North-West University in South Africa.

Defining Load Rejection

Under normal operating conditions, load following is understood as cases where the rate of change of power per minute is $<10\%$ of the maximum continuous rating (MCR). In the closed cycle, this can normally be achieved by controlling the inventory level. This implies that coolant is either extracted from or injected into the system, resulting in a related change in system pressure. This results in a change in density, which in turn results in a change in mass flow rate and, therefore, also a change in power level.

In contrast to load following, load rejection is defined as all cases where the required rate of change of power output is $>10\%$ of MCR per minute. In the case of the plant under investigation, load rejection must be achieved without allowing the plant to shut down completely. This includes an instantaneous drop in grid (export) power to any level between zero and 100% of MCR due to unforeseen circumstances. The most severe of these is when the export power drops instantaneously from 100% to zero percent of MCR. A typical scenario is in the case of an unforeseen power trip. Since normal inventory control cannot accommodate the required drop in mass flow within the short times involved, other measures to rapidly reduce the fluid power supplied to the power turbine had to be investigated.

Since the fluid power cannot be dissipated instantaneously due to inertia and compressibility effects within the system, some degree of overspeed of the power turbine and generator will always occur. The allowable degree of overspeed is specified in terms of both the absolute rotational speed measured in Hertz as well as the integrated area under the speed versus time curve measured in terms of cycles per second times time. The purpose of this study is therefore to investigate different options for controlling the degree of overspeed of the power turbine/generator assembly during full load rejection.

Control Requirements

The first step before the design of the control system could be addressed was to determine the "user requirements" for the system. From this it was specified that the load rejection control system had to:

- Enable full load rejection from 100% of MCR to zero power output to the grid within a specified time limit. Zero power output to the grid implies that the generator still supplies house load, but no grid power. For the purpose of this study, house load was set at 2.5 MW.
- Limit overspeed to within 10% of normal operating speed of 50 Hz.

Submitted to ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received December 29, 2004; final manuscript received April 27, 2006. Review conducted by Septimus van der Linden.

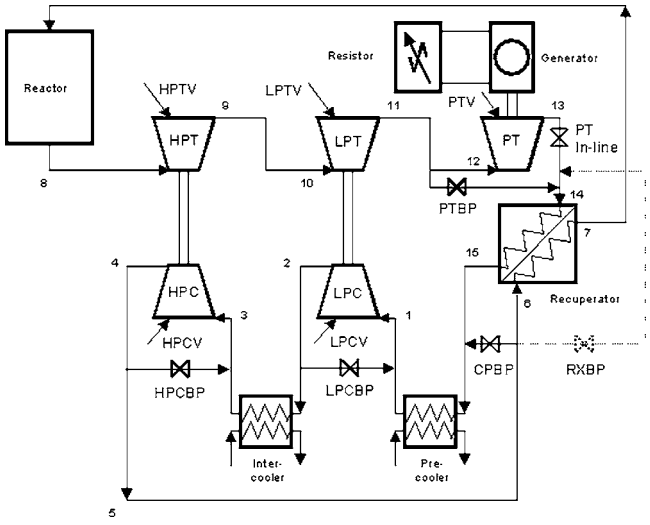


Fig. 1 Schematic representation of a high-temperature gas-cooled reactor Brayton cycle

- Limit the maximum integrated area under the overspeed versus time curve to a predetermined limit. For the purpose of this study, this was taken to be 120 Hz s.

In order to demonstrate the potential of the various options, typical values for the various requirements were chosen based on anticipated values for commercially available components. This resulted in a generator frequency of 50 Hz to correspond with the existing South African supply frequency.

Traditional Control Methods

The first step in determining a suitable control philosophy was to investigate the strategies used for controlling overspeeding of traditional power generation systems. This included both gas turbine as well as steam turbine systems.

In both types of plants, the control system typically monitors the turbine shaft speed and trips the turbine unit when the speed exceeds 110% of operating speed. Overspeed bolts are further used to mechanically trip the unit when the overspeed reaches 12.5% (112.5% of operating speed) [2].

Gas turbine systems normally measure the shaft speed with the aid of magnetic transducers placed alongside a toothed wheel. The transducers then provide an output in the form of an ac voltage with a frequency proportional to the rotational speed of the shaft. The voltage is then manipulated to indicate a speed on the control panel. If the measured speed differs from the desired speed, then the control system will adjust the fuel valve to the combustion chamber to eliminate the difference within the shortest possible time. In the case of an emergency shutdown, such as a fault on the grid, the circuit breakers and fuel valves are tripped immediately without waiting for the load to be reduced.

The thermal inertia of a reactor containing a large graphite mass implies that the heat source is subject to very slow transients. This means that this conventional option cannot be employed without shutting the system down altogether. The problems associated with restarting the reactor is also significantly more complex than the restart of a traditional gas turbine, and therefore, it is essential that the system is not allowed to shut down completely.

Traditional steam turbine power plants normally use hydraulic control systems to transmit the signals from the speed or pressure transducers to the steam control valves. On more modern units, electrohydraulic controls together with high system pressures ensure rapid response of the steam valves in case of an emergency trip (where the grid is lost). The steam valves simply shut off the steam supply to the turbine whenever the turbine exceeds a predetermined value. This is normally achieved though intercept valves placed upstream of the intermediate-pressure turbine to interrupt the flow of steam to the power turbine. It quickly reduces the fluidic power available to drive the turbine by reducing the steam flow rate.

Peet and Leung [3] and Leung and Peet [4] investigated the development of a simulation model and control system for a power turbine by-pass valve on a steam turbine plant using first principles. They identified the use of a power turbine by-pass valve as a successful method to control the open-loop steam cycle during and after load rejection.

Table 1 Different control strategies investigated

Option	Description
1	Power turbine by-pass valve together with a power turbine in-line valve
2	Variable guide vanes on all the turbines
3	Main compressor by-pass together with variable guide vanes on the turbines
4	Power turbine by-pass valve together with variable guide vanes on the HP and LP turbines
5	Main compressor by-pass valve only
6	Recuperator by-pass valve with a resistor bank
7	Main compressor by-pass valve together with variable guide vanes on the compressors
8	Main compressor by-pass valve together with local by-pass valves over the HP and LP compressors

Development and Validation of Models

The complexity associated with the thermohydraulic design of the three-shaft Brayton cycle called for the use of a wide variety of analysis techniques and simulation tools. One of the most prominent of these is the FLOWNEX thermohydraulic network simulation software developed at the Engineering Faculty of the North-West University [5]. FLOWNEX allows detailed steady-state and transient thermohydraulic simulations fully integrated with proportional, integral, and differential (PID) controller algorithms. The simulation ensures complete conservation of mass, momentum, and energy while accounting for nonideal gas behavior and compressibility effects, such as choked flow through orifices. A reactor model consisting of a point kinetic neutronics model combined with a detailed one-dimensional finite difference thermohydraulic simulation is also included [6]. Finally, the dynamics of the turbounits as well as thermal storage in heat exchangers are also accounted for.

FLOWNEX simulations are especially useful for component integration as well as for the study of complex load following and load rejection scenarios and the design of suitable controller algorithms. One of the most severe load control scenarios is that of full load rejection due to the loss of the grid power. Using FLOWNEX, an investigation was subsequently performed of a number of concepts for controlling the gas turbine power plant during and directly after load rejection. These concepts were evaluated by simulation, and the major advantages and disadvantages for each concept were identified. This paper gives a brief description of the various options and the results obtained, after which it discusses the two most promising options in more detail.

Control Strategy Options

Various strategies for controlling the closed-loop Brayton cycle during full load rejection were investigated. Although Drecher [7] looked at some of these methods, he could not realize the required control for full load rejection. This suggested that other options, including different combinations of variable guide vanes and flow valves, must be investigated. Transient simulations were performed for each concept based on the load rejection specifications mentioned earlier. The concepts investigated are listed in Table 1.

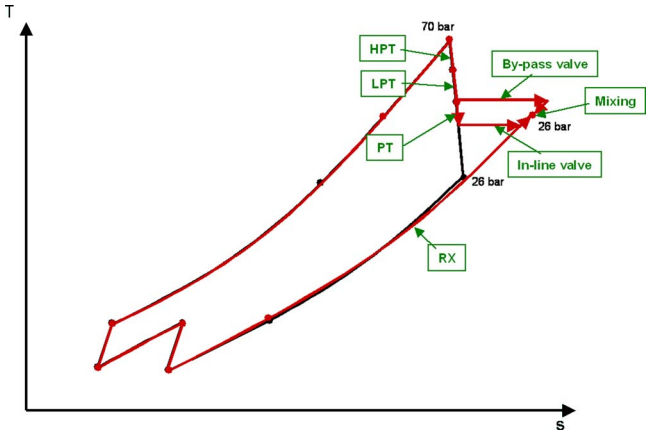


Fig. 2 *T-s* diagram—PTBP & PTIL

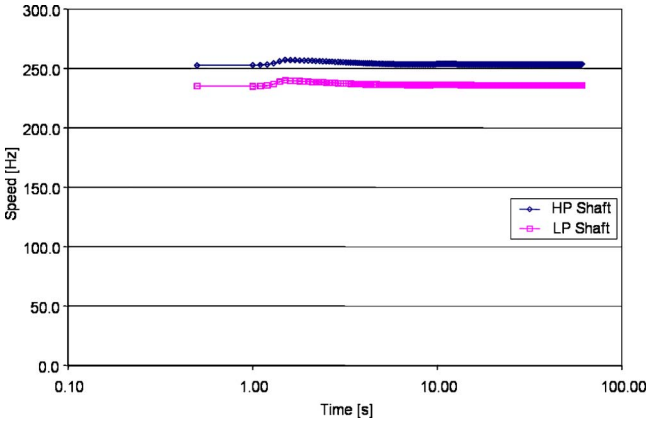


Fig. 4 Turbo unit speeds—Option 8

Option 1—Power Turbine By-Pass and Power Turbine In-Line Valves. As shown by Peet and Leung [3] and Leung and Peet [4], a power turbine by-pass valve diverting the mass flow away from the power turbine, while still maintaining the cycle resistance through an in-line valve, offers an effective control option for the traditional gas or steam turbine layout. The question subsequently arises whether this would also be true for a three-shaft closed-cycle configuration.

Objective. The objective of the power turbine by-pass (PTBP) is to divert the flow of coolant around the power turbine instead of through it. Although this also tends to reduce the overall system resistance, this can be compensated for by choking the flow through a valve placed in line with the power turbine (PTIL). The advantage is that only the mass flow through the power turbine is affected while the rest of the system remains unaffected. This enables the HP and LP turbounits to be operated at exactly the same conditions after load rejection as when generating full power.

Results. Figure 2 shows the change in system operating conditions on a *T-s* diagram for Option 1. It can be seen that, apart from the section around the power turbine, the operational conditions throughout the system remains virtually unchanged.

Figure 3 shows the variation in the system power levels. It can be seen that the shaft power lags the loss in grid power. This is simply due to the inertia of the system. However, it can be seen that it was possible to fully reject the load right down to only the house load.

With the operating conditions for the HP and LP turbounits remaining unchanged, it can be expected that the turbounit shaft speeds remain fairly constant as well. This is confirmed by the results shown in Fig. 4.

Furthermore, the small variation in compressor operating conditions has the effect that the operating points for both compressors subsequently remain sufficiently far from the surge line (see Figs. 5 and 6). The control of the system is effected by rapidly closing the power turbine in-line valve to limit the mass flow through the turbine while opening the by-pass valve to ensure constant mass flow through the rest of the system. Simulations showed that it is possible to limit the power turbine overspeed to 3.5 Hz, returning it to 50 Hz within 60 s (see Fig. 7). This translates to a cumulative overspeed of 45 Hz s, which is well within the specified limits. Figure 7 shows the resulting graph for the power turbine shaft speed.

The next important aspect to consider is the effect of the control strategy on the recuperator temperatures. From Fig. 1, it can be

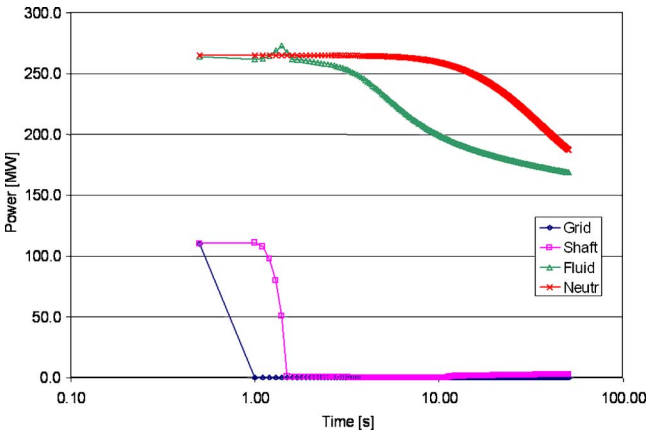


Fig. 3 Power levels—Option 8

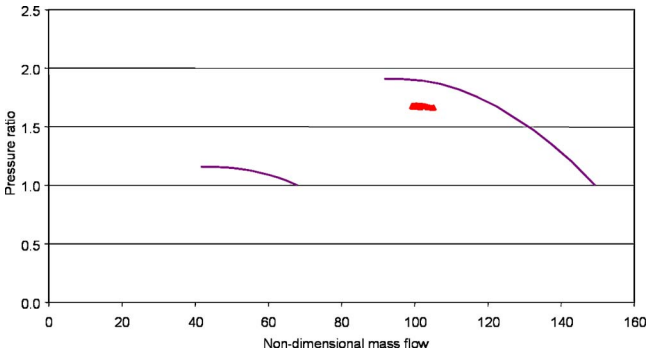


Fig. 5 HP compressor performance—Option 1

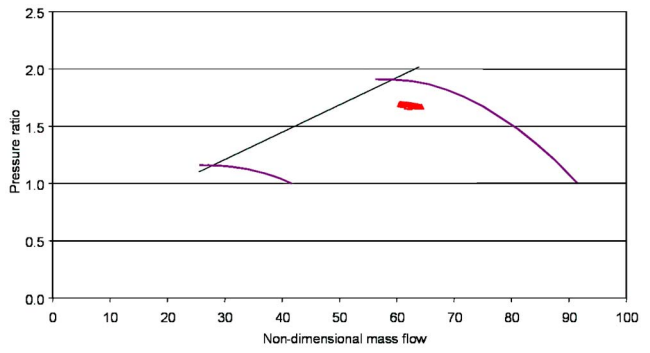


Fig. 6 LP compressor performance—Option 1

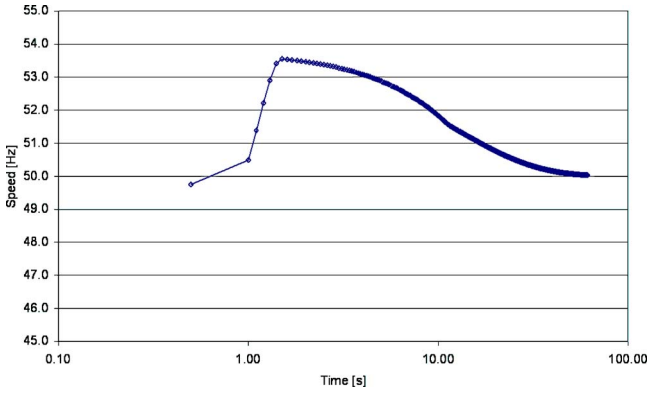


Fig. 7 Power turbine speed—Option 1

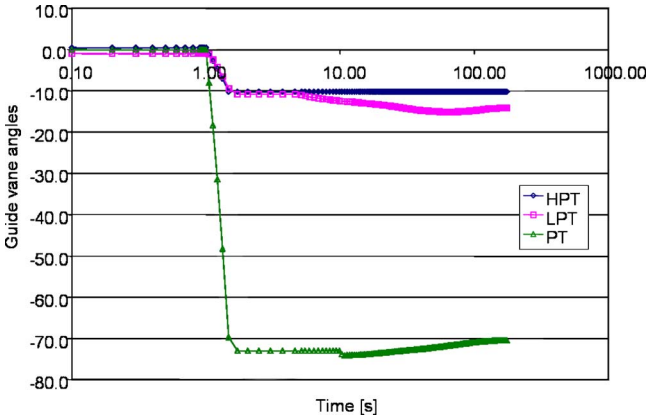


Fig. 10 Required guide vane angles—Option 2

seen that short-circuiting the power turbine results in hotter helium entering the LP side of the recuperator. This is confirmed by the recuperator temperature results shown in Fig. 8. This sudden change should be acceptable as long as a reasonably homogeneous temperature distribution is ensured at the inlet to the heat exchanger. The delay in the temperature increase of the HP coolant is due to the thermal mass of the recuperator.

A final important aspect to consider is the effect of the increased recuperator exit temperature on the reactor temperatures. From Fig. 9, it can be seen that although the reactor inlet temperature rises along with the recuperator exit temperature, the pebble temperature and reactor outlet temperature remains constant. Option 1 therefore offers an elegant control strategy capable of adequately controlling the system during full load rejection.

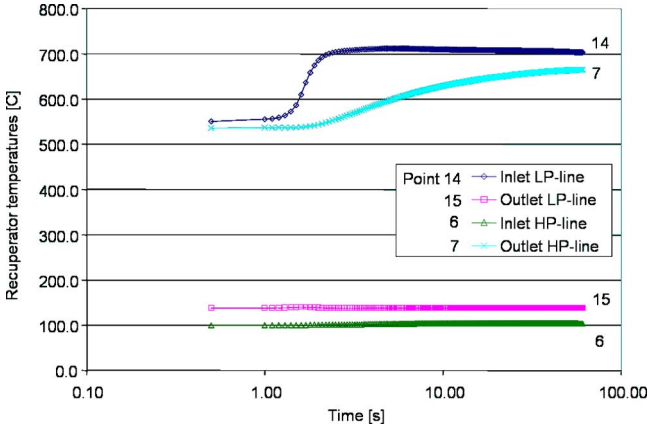


Fig. 8 Recuperator temperatures—Option 1

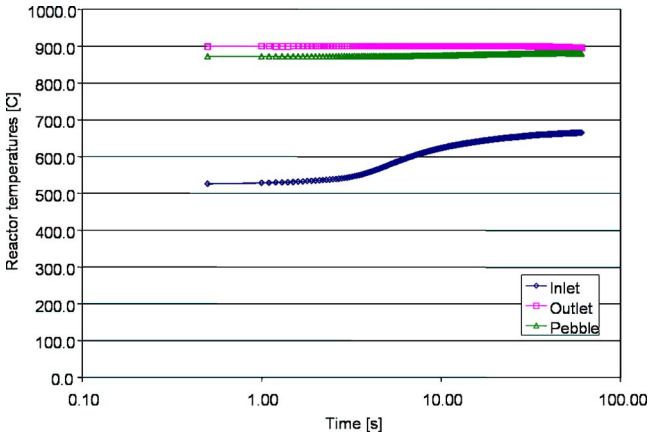


Fig. 9 Reactor temperatures—Option 1

Advantages/Disadvantages. The first major advantage of Option 1 is the absence of variable inlet guide vanes on the turbines operating in the high-temperature helium environment. The system can also be returned to stable operating conditions (at a power turbine speed of 50 Hz) well within the specified limits. However, the configuration does have a number of limitations. The first is that it requires intricate, high-temperature helium valves operating at temperatures of around 700°C. Furthermore, it requires a mixing chamber for mixing the warmer fluid stream by-passing the power turbine with the cooler stream flowing through and exiting the power turbine. This is to satisfy the requirement of a uniform temperature distribution across the recuperator in order to limit the effect of thermal shocks on the recuperator construction.

Option 2—Variable Guide Vanes on All the Turbines. In Option 2 all three turbines (HPTV, LPTV, and PTV) were assumed to be fitted with variable guide vanes (initially open). The aim was to use the guide vanes to drastically reduce the efficiencies of the turbines during load rejection. Assuming a relatively constant mass flow, this increases the pressure ratio required for the turbounits to drive the compressors. This reduces the pressure ratio over the power turbine and, therefore, also the power available to drive the generator. The power turbine guide vanes are then adjusted in a similar fashion to further reduce the available work from the power turbine to acceptable levels before the inertia of the generator assembly allows the shaft speed to increase.

A controller subsequently monitors the shaft speed closely and adjusts the guide vanes by means of a PI controller, maintaining the shaft speed at close to 50 Hz. The inlet guide vanes of the HP and LP turbines are also adjusted in order to maintain system temperature conditions similar to that experienced during full load operation.

Detailed simulations showed that it is possible to successfully control the cycle during a full load rejection to 0 MW grid power. This was achieved by a 75 deg adjustment in the power turbine inlet guide vanes. Although the delayed reaction between the loss of the grid power and the first response resulted in an increase in shaft speed, it was limited to ~53 Hz with an integrated overspeed of only 40 Hz s. The speed of the compressors was furthermore also stabilized at speeds close to that experienced during full load operation.

The main limitation of Option 2 is the intricate design and manufacture of controllable guide vanes capable of reliably adjusting through 75 deg within the required time limit in the high temperature helium environment (Fig. 10).

Option 3—Main Compressor By-Pass With Variable Guide Vanes on All the Turbines. Option 3 investigated the use of a main compressor by-pass (CPBP) to short-circuit the turbines, together with variable guide vanes on all the turbines (HPTV, LPTV, and PTV). The aim of the compressor by-pass was to reduce the mass flow through the turbines while keeping the mass flow through the compressors constant. This leaves the power input to the compressors unchanged. With this power provided by the HP and LP turbines, the fluidic power available to the power turbine is reduced.

Using a combination of the by-pass valve and guide vanes it was possible to successfully control the system during load rejection. The reduced mass flow through the turbines, however, resulted in an increase in the minimum self-sustaining power level. This is due to the fact that if the output power level is reduced even further, the HP and LP turbines would not be able to supply sufficient power to drive the compressors. It subsequently resulted in the requirement for a resistor bank with a continuous rating of 5 MW to dissipate the excess output power in order to ensure self-sustainable operation. However, it was possible to limit overspeed to below 3 Hz and the integrated overspeed to 15.3 Hz s. Although the need for guide vanes was not eliminated, the maximum adjustment angle was reduced to <40 deg. The advantage of the smaller variation in guide vane angles should, however, be weighted against the need for a resistor bank.

Option 4—Variable Guide Vanes on the HP And LP Turbines Together With a Power Turbine By-Pass Valve. Option 4 investigated the possibility of completely eliminating the power turbine guide vanes. Although variable guide vanes were still used on the HP and LP turbines (HPTV and LPTV, respectively), the power turbine guide vanes were replaced by a power turbine by-pass valve (PTBP). With no guide vanes to adjust pressure ratio, the intention was to use the PTBP to divert the flow around the power turbine. This would drastically reduce the mass flow through the power turbine, resulting in a greatly reduced power output.

Although the configuration initially seemed to have potential, FLOWNEX simulations showed that opening the power turbine by-pass valve reduced the system resistance and therefore increased the mass flow throughout the rest of the cycle. This resulted in the opposite of the intended reaction with the power turbine overspeeding and the turbounit shaft speeds increasing significantly. Even after adjusting the guide vanes significantly more than in Option 2, turbine overspeeding could not be prevented without the system shutting down. The option was subsequently discarded as a possible control option.

Option 5—Compressor By-Pass Valve Only. In a further attempt to eliminate the use of turbine guide vanes, the possibility of only using of a compressor by-pass valve was investigated. As with Option 3, the aim was to quickly reduce the mass flow through the turbines by momentarily short-circuiting the inlet and outlet of the compressor section. The effect would be that the mass flows through the turbines are substantially reduced while keeping the mass flows through the compressors the same. This again leaves the power input to the compressors unchanged. With this power provided by the HP and LP turbines, the fluidic power available to the power turbine is reduced. This leaves significantly less fluidic power available to the power turbine. Major advantages of this configuration were that no variable guide vanes were required and existing valve technology could be used for the CPBP due to the lower gas temperatures.

Simulations using only a CPBP showed that it was not possible to obtain stable operation or prevent the system from shutting down during load rejection. However, by adding a resistor bank of 50 MW, it was possible to obtain stable operation by utilizing the full resistance capacity for 2 s, after which it was reduced 10 MW. Overspeed was once again limited to <53 Hz, with the integrated overspeed being limited to a mere 2.2 Hz s. The turbounit shaft speeds were also reduced significantly and stabilized at the reduced power levels. However, apart from the resistor bank, a major drawback of this configuration is that the LP compressor tends to run into surge during the load rejection. It furthermore also resulted in a sharp increase in the recuperator LP inlet temperature as shown in Fig. 11.

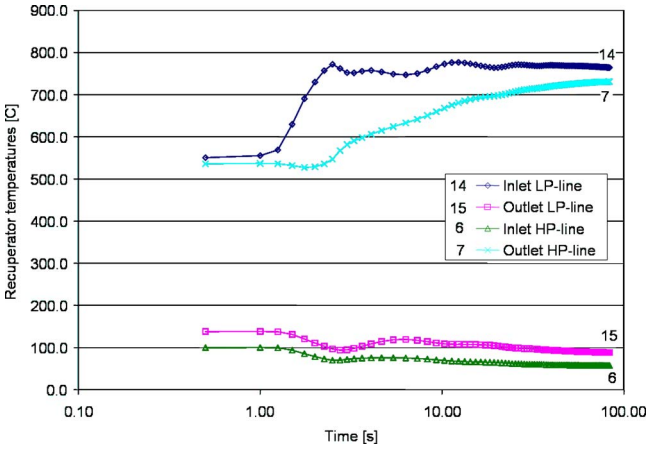


Fig. 11 Variation in recuperator temperatures—Option 5

Option 6—Recuperator By-Pass Valve With a Resistor Bank. The aim of Option 6 was to effect the same control as Option 5, but introducing cooler helium from the manifold in front of the recuperator in order to limit the recuperator inlet temperature on the LP side. This was done by adding a recuperator by-pass (RXBP) valve. However, this option did not pose any advantage over Option 5. On the contrary, it resulted in sharply fluctuating temperatures at the recuperator LP inlet (see Fig. 12) due to the varying amount of cooler helium being passed through the adjustable by-pass valve. The manufacturers further confirmed that this poses more of a problem than the sudden increase in temperature observed with Option 4. It also requires an intricate mixing chamber specially designed to ensure homogeneous mixing of the hot stream from the turbines and the "cold" by-pass stream before it enters the recuperator.

Option 7—Compressor By-Pass Together With Variable Guide Vanes on the HP and LP Compressors. Apart from the potential compressor surging, Option 5 offered good potential for controlling the system to run stable during load rejection. One traditional method of controlling surge is by using variable guide vanes on the compressor. It therefore made good sense to inves-

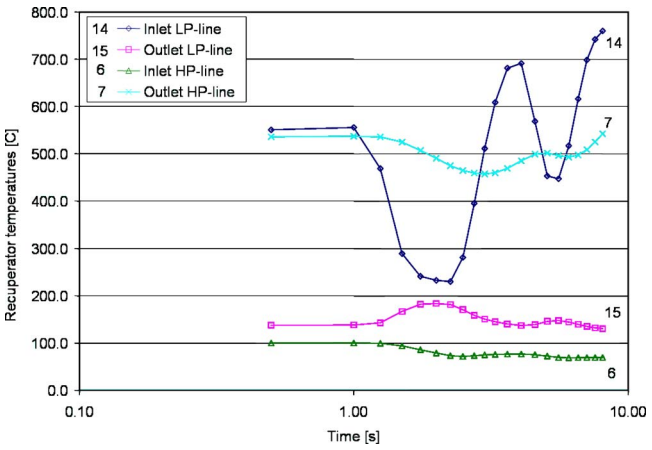


Fig. 12 Variation in recuperator inlet temperature—Option 6

tigate the possibility of using Option 5 together with variable compressor guide vanes (HPCV and LPCV) to effect the required control.

Simulations proved that it was possible to control the system without the compressor surging while turbine over-speed was also limited to within the specified ranges. A major advantage over Options 2–4 is that the variable compressor guide vanes operate in a relatively benign environment compared to variable turbine guide vanes. Load rejection was achieved down to 10 MW, requiring a resistor bank with a similar continuous rating. All other results were similar to that obtained with Option 5.

Option 8—Compressor By-Pass and Local By-Pass Valves Over the HP and LP Compressors. Although offering the potential for controlling the system during load rejection, variable guide vanes remain an intricate and rather expensive option. Based on the results obtained with Options 5 and 7, other alternatives eliminating the need for variable guide vanes and preventing the compressors from surging were investigated. As surge is, inter alia, caused by a too low mass flow, the possibility of adding a local by-pass valve in order to increase the mass flow through the relevant compressor was evaluated.

As in Option 5, this option also utilizes a compressor by-pass valve (CPBP) to quickly reduce the mass flow through the turbines by momentarily short-circuiting the compressor inlet and outlet. The valve is then closed again to maintain stable operation. A resistor bank is employed to dissipate excess energy and perform fine control. It further has a local by-pass valve (HPCBP and LPCBP) fitted to each compressor while the compressors and turbines are all equipped with fixed guide vanes. Simulations showed that Option 8 offers good potential for ensuring load rejection control, and it is therefore discussed in more detail.

Objective. The objective of Option 8 is to reduce the mass flow through the turbines almost instantaneously, thus reducing the mass flow available to generate power. The aim is to maintain a reduced mass flow through the turbines while keeping the mass flow through the compressor close to constant even after load rejection, thus preventing the compressors from running into surge. This is supported further by increasing the mass flow through the compressors by means of a local by-pass valve across both the HP and LP compressors.

Results. In order to control the system and prevent overspeed, it is essential to "remove" the source of energy to the turbine. One such a method is opening the compressor by-pass valve quickly, thus reducing the maximum cycle pressure while increasing the minimum cycle pressure. The reduced pressure ratio results in a reduced fluidic power to the turbine, which in turn results in a significantly lower power output from the power turbine. Simulations showed that the system could be successfully controlled down to a minimum power level of 10 MW. This load then has to be dissipated via a resistor bank with a similar rating. Figure 13 shows the resulting *T-s* diagram, whereas Fig. 14 shows the resulting power levels during load rejection.

Figure 15 shows that a substantial drop in the HP and LP turbo-unit shaft speeds was experienced due to the reduced mass flow. However, the units eventually stabilized, sustaining the system at the lower power level. As the compressor shaft speeds reduce, the compressor operating points tend to move toward the surge line. Opening the local by-pass valves now controls the mass flow through the compressors individually, shifting the compressor operating point away from the surge line, thus enabling stable operation of the system (confirmed in Figs. 16 and 17).

In order to ensure the shortest reaction time for the power turbine the main compressor by-pass is opened rapidly with the local by-pass valves being allowed to adapt according to the compressor reaction. After a predetermined time the main valve is closed and the system is controlled by the local by-pass valves together with the main by-pass. Figure 18 shows the respective valve opening diameters. With this control strategy it was possible to limit the power turbine overspeed to <3 Hz (see Fig. 19) and a cumulative overspeed of <10.0 Hz s.

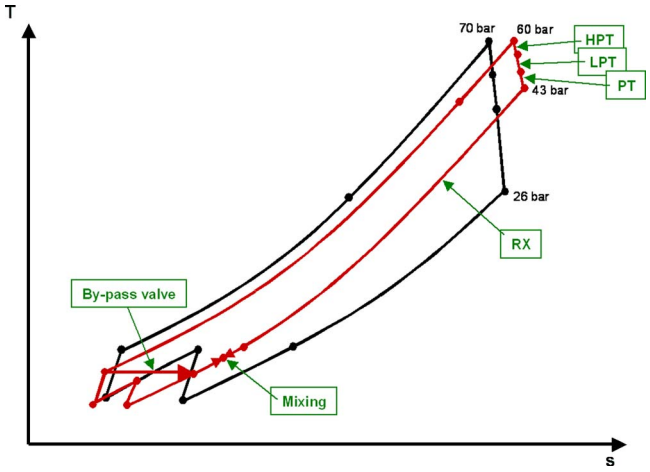


Fig. 13 *T-s* diagram—CPBP with HPBP and LPBP

From Fig. 20, it can be seen that the temperature of the coolant entering the LP side of the recuperator is significantly higher after load rejection than under normal operating conditions. As mentioned before, this should not pose a problem as long as it is sufficiently mixed to ensure homogeneous temperature distribution over the inlet area.

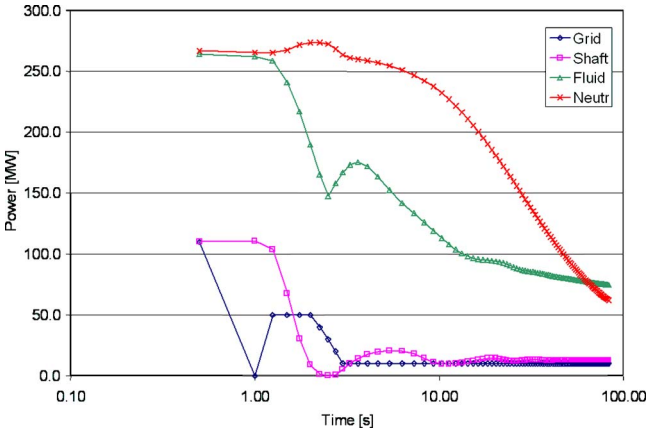


Fig. 14 System power levels—Option 7

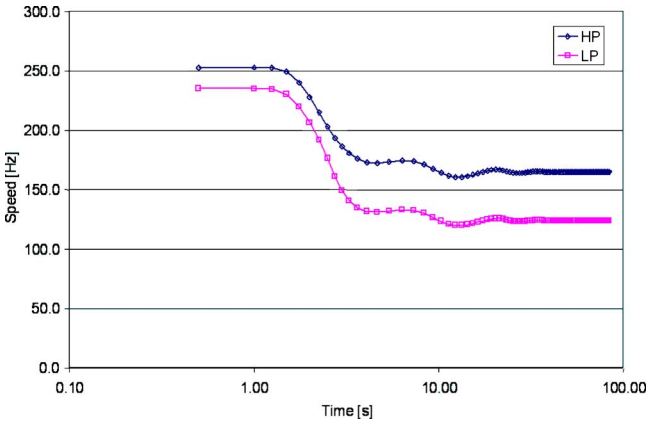


Fig. 15 Turbo unit shaft speeds—Option 7

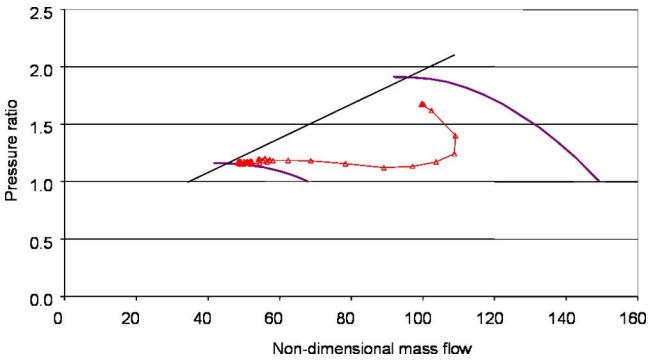


Fig. 16 LPC performance—local by-pass valves

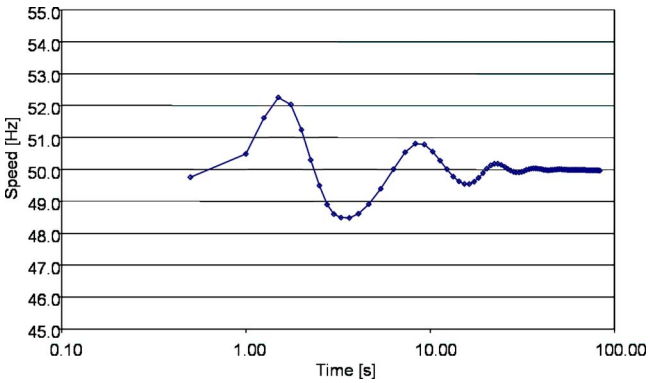


Fig. 19 Power turbine speed—Option 7

Advantages/Disadvantages. The first main advantage of this option is the possibility of using existing valve technology due to the significantly lower valve operating temperatures. Furthermore, the option also eliminates the need for variable guide vanes on the turbomachines with the system returning to stable operating conditions (at a power turbine speed of 50 Hz) well within the limits specified earlier. Only two limitations of this option could be identified thus far. The first is that it is not possible to reject load to zero power level, thus requiring a resistor bank with a continuous rating of 10 MW. However, in order to limit overspeeding during the transient, the resistor bank requires a minimum capacity of 50 MW.

Although initial simulations showed that Option 8 has good potential as a control strategy for a three-shaft system, further investigation of the option will have to be performed to confirm its suitability.

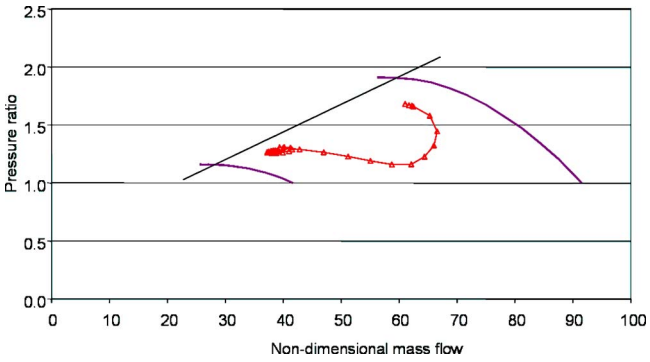


Fig. 17 HPC performance—Local HP valve closed, but with the required local LP valve adjusted

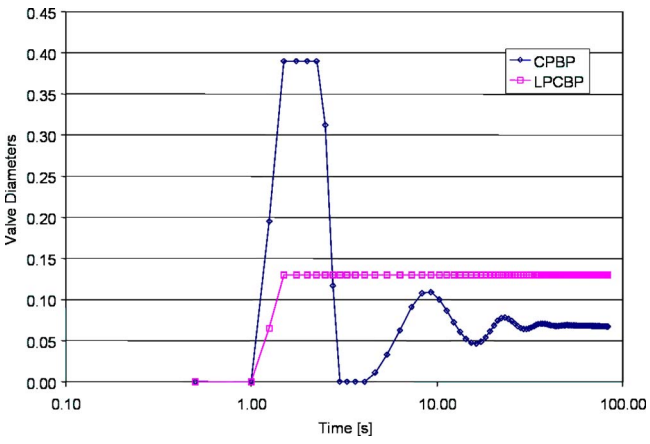


Fig. 18 Control strategy—valve opening—Option 8

Conclusions

A number of different control strategies were investigated for controlling a three-shaft closed-loop gas turbine power plant cycle during load rejection. Table 2 provides a summary of the most important advantages and disadvantages of each. Values stated are purely indicative of the relative performance between the various options and are based on anticipated values for commercially available components.

The results showed that Option 1 offers an elegant method for controlling the system during load rejection using a power turbine in-line and by-pass valve. Although it offers the opportunity of full load rejection, it will require valves operating in a high-temperature helium environment. Another disadvantage is that an intricate mixing chamber will be required to ensure a homogeneous inlet temperature distribution to the inlet of the recuperator.

The most promising alternative to Option 1 was found to be a compressor by-pass valve and local by-pass valves over the HP and LP compressors as contained in Option 8. A major advantage of this option is that it can be based on existing valve technology since the valve operating temperature is $\sim 150^{\circ}$C. Simulations using FLOWNEX showed that it is indeed possible to successfully control the system during load rejection. However, full load rejection cannot be achieved and a resistor bank with a short-term power dissipation capacity of 50 MW is required to limit overspeeding to within the specified limits. A minimum continuous power dissipation capacity of 10 MW is required to maintain the minimum level of power output over a prolonged period. How-

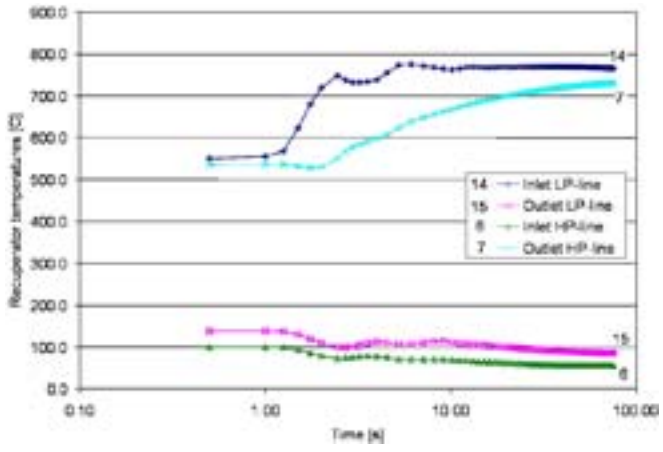


Fig. 20 Recuperator temperatures—Option 8

Table 2 Summary of control options

Option	Advantages	Disadvantages
1—PTBP and PTIL valves	No variable turbine guide vanes Full load rejection without resistor bank Stable operations regained within specified limits	High-temperature helium valves Intricate mixing chamber required
2—HPTV, LPTV, and PTV	Full load rejection without resistor bank	Excessive blade angles around 75 deg High-temperature variable turbine guide vanes
3—CPBP with HPTV, LPTV, and PTV	Load rejection possible Blade angles <40 deg Excellent overspeed characteristics	5 MW resistor bank required High-temperature variable turbine guide vanes
4—HPTV and LPTV with PTBP	—	Not possible to regain stable operation without shutting down
5—CPBP only	No variable turbine guide vanes Existing valve technology Impressive overspeed characteristics	50 MW transient resistor bank to improve overspeed characteristics 10 MW continuous rating resistor bank required LP compressor runs into surge
6—RXBP with resistor bank	No real advantage over Option 5	Nonhomogenous thermal distribution to recuperator Intricate mixing chamber
7—CPBP, HPCV, and LPCV	Friendly operating environment for guide vanes	10 MW resistor bank Costly compressor guide vanes
8—CPBP, HPCBP, and LPCBP	Existing valve technology No variable guide vanes required	50 MW transient resistor bank to improve overspeed characteristics 10 MW continuous rating resistor bank

ever, the presence of the resistor bank has the added advantage that it may also be used for finer power output control under normal operating conditions. Although the technology involved is already available, the cost of such a resistor bank could be prohibitive.

Based on the above information, Option 8 could offer a lower risk short-term solution. However, because of the elegance from a thermodynamic point of view, the development of the valve technology required for Option 1 may be worth investigating as a longer-term solution.

Acknowledgment

This study formed part of the initial design investigations for the PBMM, which was conducted in association with M-Tech Industrial (Pty) Ltd. and funded by PBMR (Pty) Ltd.

Nomenclature

ESKOM = South African electrical utility company
FLOWNEX = thermohydraulic systems CFD simulation software developed at North-West University, South Africa, www.flownex.com
HP = high-pressure
LP = low-pressure
MCR = maximum continuous rating
PBMM = pebble bed micromodel
PBMR = pebble bed modular reactor
PID = proportional, integrator, and differential controller

References

[1] Greyvenstein, G. P., and Rousseau, P. G., 2003, "Design of a Physical Model of the PBMR With the Aid of Flownet," Nucl. Eng. Des., **222**, pp. 203–213.
[2] Kiameh, P., 2002, *Power Generation Handbook*, McGraw-Hill, New York.
[3] Peet, W. J., and Leung, T., 1995, "Development and Application of a Dynamic Simulation Model for a Drum Type Boiler With Turbine Bypass System," International Power Engineering Conference, Singapore.
[4] Leung, T., and Peet, W. J., 1995, "Control System Retrofits to Improve Plant Efficiency," Canadian Electrical Association, Thermal Generating Station Construction and Commissioning Session, Vancouver, Canada.
[5] Greyvenstein, G. P., Van Ravenswaay, J. P., and Rousseau, P. G., 2002, "Dynamic Modeling of Heat, Mass and Momentum Transfer in the Pebble Bed Modular Reactor," 1st International Conference on Heat Transfer, Fluid Mechanics, and Thermodynamics, Kruger Park, South Africa, April, 8–10.
[6] Rousseau, P. G., and Greyvenstein, G. P., 2002, "One-Dimensional Reactor Model for the Integrated Simulation of the PBMR Power Plant," 1st International Conference on Heat Transfer, Fluid Mechanics, and Thermodynamics, Kruger Park, South Africa, April, 8–10.
[7] Decher, R., 1994, *Energy Conversion—Systems, Flow Physics and Engineering*, Oxford University Press, London.

Reduced-Order Modeling and Wavelet Analysis of Turbofan Engine Structural Response due to Foreign Object Damage (FOD) Events

James A. Turso[1]
QSS Group, Inc.,
21000 Brookpark Road,
Cleveland, OH 44135

Charles Lawrence
NASA Glenn Research Center,
21000 Brookpark Road,
Cleveland, OH 44135

Jonathan S. Litt
U.S. Army Research Laboratory,
Glenn Research Center,
21000 Brookpark Road,
Cleveland, OH 44135

The development of a wavelet-based feature extraction technique specifically targeting FOD-event induced vibration signal changes in gas turbine engines is described. The technique performs wavelet analysis of accelerometer signals from specified locations on the engine and is shown to be robust in the presence of significant process and sensor noise. It is envisioned that the technique will be combined with Kalman filter thermal/health parameter estimation for FOD-event detection via information fusion from these (and perhaps other) sources. Due to the lack of high-frequency FOD-event test data in the open literature, a reduced-order turbofan structural model (ROM) was synthesized from a finite-element model modal analysis to support the investigation. In addition to providing test data for algorithm development, the ROM is used to determine the optimal sensor location for FOD-event detection. In the presence of significant noise, precise location of the FOD event in time was obtained using the developed wavelet-based feature. [DOI: 10.1115/1.2718230]

1 Introduction

Ingestion of foreign objects by a turbofan engine, particularly birds and ice, is one of the leading causes of uncontained rotor events in commercial jet transport. There are numerous cases of bird ingestion contributing to accidents, some of them fatal [1]. Specific procedures have been developed for situations where ingestion is suspected, and forensic analysis has repeatedly shown that they would have been the appropriate actions to take in cases where rejected takeoffs motivated by bird ingestion resulted in accidents. These are the worst situations; in the vast majority of cases, bird ingestion does not affect the safe outcome of a flight and may, in fact, go unnoticed by the flight crew. However ingestion does pose a risk of foreign object damage (FOD), and thus is important to identify, if possible. When a multi-engine aircraft flies through a flock of birds, potentially damaging more than one engine, the pilot needs to understand the status since his/her resulting actions may be different depending on the number of engines involved. Even in a case where there is no apparent damage to an engine as observed from the cockpit, latent effects (e.g., cracks that can be propagated by high cycle fatigue) may be present that require maintenance action [2].

The critical consequence of foreign object ingestion is engine surge, potentially resulting in the loss of power. The flight crew can recognize foreign object ingestion through a combination of instrument readings and sensory cues. These include such symptoms as a thud or bang, a fire warning, a visible flame coming out of the engine, vibration, yaw of the airplane caused by thrust imbalance, high exhaust gas temperature, change in the spool speeds, smoke/odor in cabin bleed air, and engine pressure ratio change. It is important to note that for impact-type FOD (due to ice, birds, runway debris, etc.), the damage is primarily to the fan and front part of the engine, with the extent of the damage determined by the geometry, angle of impact, hardness, relative speed, etc., of the object.

One key reason for detecting FOD automatically is that most such events occur close to the ground when the flight crew is occupied flying the plane. Only after the crew stabilizes the aircraft at a safe altitude should they take action on the engine. Once the aircraft is stable, the reduced workload in the cockpit environment provides better circumstances under which the crew can analyze the situation, and they would benefit greatly by having full knowledge of the engines involved and the likelihood of the event based on data.

Given that many of the potential symptoms of FOD are not unique to that type event, an automatic system for FOD detection should take information from multiple sources to provide confidence in its diagnosis. Just like a pilot does, this system must fuse information in a way that provides a measure of the likelihood of foreign object ingestion in order to determine any corrective action. Additionally, since several of the potential symptoms are described in terms of human senses, alternate information sources need to be developed. These sources should utilize the standard engine sensor suite (or a very similar suite) to address issues such as certification and retrofit. Naturally processing requirements may vary from current systems simply due to the fact that additional on-line signal processing would be required, but the impact would be minimal since such a system would not be flight critical.

For approximately two decades techniques based on the Kalman filter algorithm have been applied to turbofan engine diagnostics [3,4]. Specifically, *analytical* (or *virtual*) *measurements* provided by Kalman filter estimators have been used for detection of engine degradation via estimation of a set of health parameters, which are generally not measurable themselves (e.g., compressor efficiency) and are calculated via knowledge of measurable quantities. The degradation monitored may be gradual in nature, e.g., worn components that increase internal clearances, resulting in decreased component efficiency. It may also be abrupt as is the case when foreign objects are ingested by the engine. Changes in component efficiencies, high and low spool speeds, and thermo-

[1]Current affiliation: Northrop Grumman Ship Systems, Advanced Capabilities Group, 6608 Sunscope Drive, Ocean Springs, MS 39564.

Submitted to ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received May 21, 2004; final manuscript received November 17, 2006. Review conducted by Marc P. Mignolet.

dynamic parameters have been determined/observed during FOD events [3,4]. In some cases, these parameter changes alone may not be conclusive proof that a FOD event has occurred. Absent from these investigations is the incorporation of structural vibration signals as a means to aid in positive identification of a FOD event.

To successfully apply available sensor fusion techniques, the same event should be detected using sensors that have significantly different physical characteristics (e.g., thermocouples compared to accelerometers) and rely on measuring physically different parameters (e.g., temperature and acceleration) [5,6]. It would appear that fusing health parameter estimates with structural response information acquired during a FOD event could provide conclusive evidence that a FOD event had occurred. Furthermore, fusion of virtual measurements and vibration signal features may enable discrimination between a normal aircraft maneuver/operating condition and a more severe FOD event even though both events appear to have similar looking health parameter data profiles (due to inherent limitations of the diagnostic model used). Similarly, vibration signals alone may possess FOD-like characteristics depending on the maneuver, e.g., a rapid change in thrust demanded from the engine would result in transmission of temporary impulse-like force imbalances through the engine structure. The combination of traditional health parameter estimation and of state-of-the-art signal processing techniques applied to structural vibration signals (e.g., wavelet analysis [7,8]) could provide the "finger print" necessary to positively identify a FOD event.

This paper describes the development of a wavelet-based feature extraction technique specifically targeting FOD-event induced vibration signal changes in gas turbine engines. It is envisioned that the wavelet-based feature will be fused with measurements or estimates of engine thermal parameters, themselves having FOD-specific response characteristics, to provide a high-confidence FOD event diagnosis. Successful automated identification of FOD events would be especially useful in unmanned (autonomous) vehicle applications, where the absence of the pilot precludes the use of human monitoring of engine conditions.

To the authors' knowledge, engine-specific vibration signatures associated with actual FOD events, e.g., rotor frequencies excited due to foreign object impact on the fan disk, are not available in the open literature. Potential sources of high-frequency structural data, specifically high-fidelity finite-element models of aircraft engines, do not easily lend themselves to sensor placement and diagnostic techniques based on well-established linear system theory. As well, these models are typically computation-time intensive due to the vast amounts of time-dependent spatial data they produce and do not lend themselves for direct application to the initial stages of control and diagnostic development and testing. To facilitate wavelet-based feature development, a reduced order model is synthesized to provide the structural response signals indicative of a FOD event. The low-order structural model is also used to determine the optimal accelerometer locations for feature extraction. This requires a linear model in state-space form to facilitate the calculation of an Observability Grammian matrix [9] that would determine whether a mode excited by a FOD event could be detected by a particular sensor arrangement.

2 Development of Reduced-Order Turbofan Rotor Structural Models

Finite-element structural analysis codes provide invaluable information concerning structural deformation, stress intensities, and modal behavior. However, finite element codes tend to be cumbersome and time consuming to work with when used to design and test control and diagnostic system algorithms, especially when performing multiple design iterations. State-space models, on the other hand, are the form of choice for many signal processing and multiple input multiple output controller applications. A method for reduced-order modeling of turbofan rotor structures is presented which uses output from a finite-element code in the form of natural frequencies and eigenvectors to create a state-space model suitable for use in CDS development and testing. The resulting state-space model can easily be ported to control system development software such as the MATLAB/SIMULINK® suite of software tools [10]. The following presents the reduced-order model development process as applied to the modeling of a generic turbofan rotor, ultimately used for generating a simulated FOD event structural response.

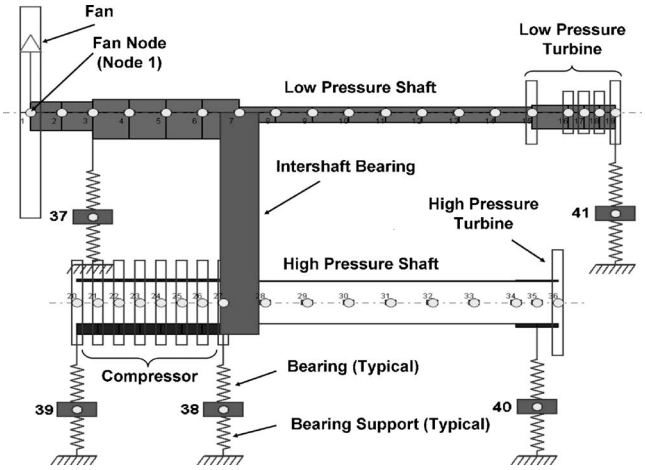


Fig. 1 DYROBES FEM schematic of turbofan rotor-bearing system

For the purpose of developing reduced-order structural models for application to CDS development, a dynamic model based on the dominant modes of vibration of a turbofan rotor-bearing system is considered adequate. The modal behavior characterized by the model should be representative of that presented in the literature for turbofan engines [11,12]. For the present study a specialized rotor dynamics finite element code, DYROBES [13], is used to simulate the structural response resulting from unbalance and foreign object damage. This code is suitable for the present study since it has the capability to model critical features of an engine system and can sufficiently simulate engine dynamic response, particularly the response at sensor locations. Being a special-purpose finite-element code, i.e., designed specifically for rotor dynamic analysis and users with a basic knowledge of FE methods, it is also relatively straightforward to use.

The structural model used for this study was derived from a generic engine model created to develop an industry standard for engine-airframe structural analysis. The model is of a dual spool rotor mounted on flexible supports (Fig. 1). The flexible supports are characterized to simulate the static engine structure and wing.

The first spool is representative of the fan and low-pressure turbine stages while the second spool represents the high-pressure compressor and turbine. An interstage bearing connects the two spools together. The low-pressure spool is 112 in. (284 cm) long and consists of 18 shaft finite elements. A large disk is located at the left end to represent the fan and several smaller disks are located on the right end for the low-pressure turbine. The entire shaft weighs 2494 lbs (1134 kg). The high-pressure spool is 80 in. (203 cm) long and consists of 16 shaft finite elements. There are eight disks on the shaft's left end representative of a compressor. One disk is located on the shaft's right end representative of the high-pressure turbine. The high-pressure shaft weighs 1280 lbs (582 kg). Each shaft is supported on bearings as shown in the figure. The bearings in turn are supported on flexible supports representative of the support flexibility provided by the static engine structure. The bearing and support stiffnesses are assumed to be 1.0×10^6 and 1.0×10^3 lb/in. (1.8×10^5 and 1.8×10^2 kg/cm), respectively.

The transient response from rotor imbalance and FOD of the

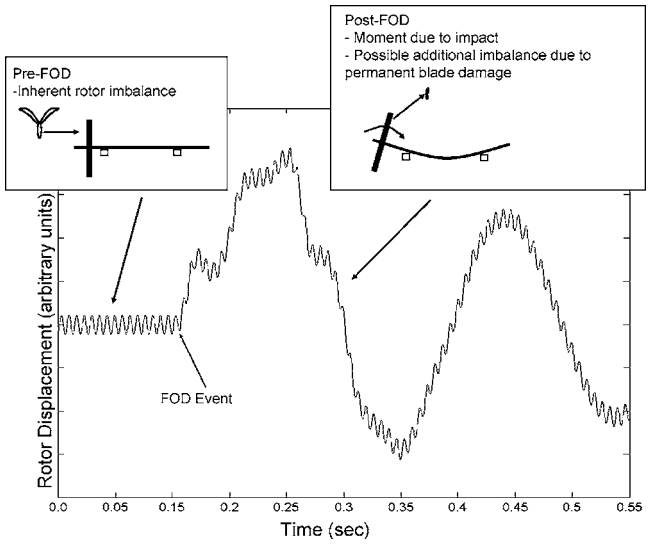


Fig. 2 Example FOD-event FEM model vibration response characteristics. Calculated displacements are functions of the imbalance magnitude and foreign object impact characteristics (e.g., relative speed, size, etc.).

turbofan DYROBES model is shown in Fig. 2. A typical rotor vibration sensor signal measured during a FOD event (with a sensor located at one or more of the bearing locations shown in Fig. 1) would consist of the response of a dominant low-frequency lateral mode excited from an impulsive moment at the fan disk/rotor shaft interface (about the *X* axis in Fig. 3) due to the initial impact of the foreign object upon the fan disk. Higher frequency lateral modes would also be excited due to the high-frequency content of the foreign object impact and would appear in the sensor signal immediately after the event. Possible longer lasting effects due to resulting permanent blade damage, i.e., an imbalance force occurring at a frequency corresponding to once per revolution of the rotor, may also appear. Rotating equipment tends to have a relatively small degree of inherent imbalance after manufacturing [11], which will also be included in the reduced order structural model for a "steady state" response. To the authors' knowledge, high-frequency accelerometer data from actual turbofan engine FOD events does not appear in the open literature, thus the hypothesized sequence of events is predominantly based on the authors' experience and intuition. The following section describes the method used to develop a reduced order model (ROM) that closely approximates the actual rotor response but is more amenable to real-time assessment of feature extraction techniques used for identifying FOD.

2.1 Dynamic Representation of Rotor Bearing Systems in the State Space. Dynamic systems may be represented as transfer functions or in state-space [14]. For CDS development the state-space form is often preferred and will be the form used for the reduced order model synthesized from DYROBES FEM modal analysis data. For a multiple degree of freedom translational system the equation of motion in physical coordinates is

$$[M]\{\ddot{q}\} + [B]\{\dot{q}\} + [K]\{q\} = \{F\} \tag{2.1}$$

where M is the system mass matrix, B is the damping matrix, and K is the system stiffness.

The dynamics of rotating systems are complicated by the presence of gyroscopic effects [11,15]. With gyroscopic effects included Eq. (2.1) becomes

$$[M]\{\ddot{q}\} + [B + G]\{\dot{q}\} + [K]\{q\} = \{F\} \tag{2.2}$$

Unlike a nonrotating system, the angular velocity-related coefficient matrix G in Eq. (2.2) is a function of the shaft rotational speed, Ω, i.e., gyroscopic effects result in the rotor's lateral natural frequencies and mode shapes being a function of Ω [11].

In the present work, an approximate modal model of the rotor was developed by using the eigenvalues corresponding to the forward modes calculated by DYROBES at a specified rotor speed and treating them as if they were the normal modes of a nonrotating structure. In this fashion the gyroscopic effects are approximately included in the modal basis without the inclusion of a gyroscopic matrix, with the resulting model being accurate only at the speci-

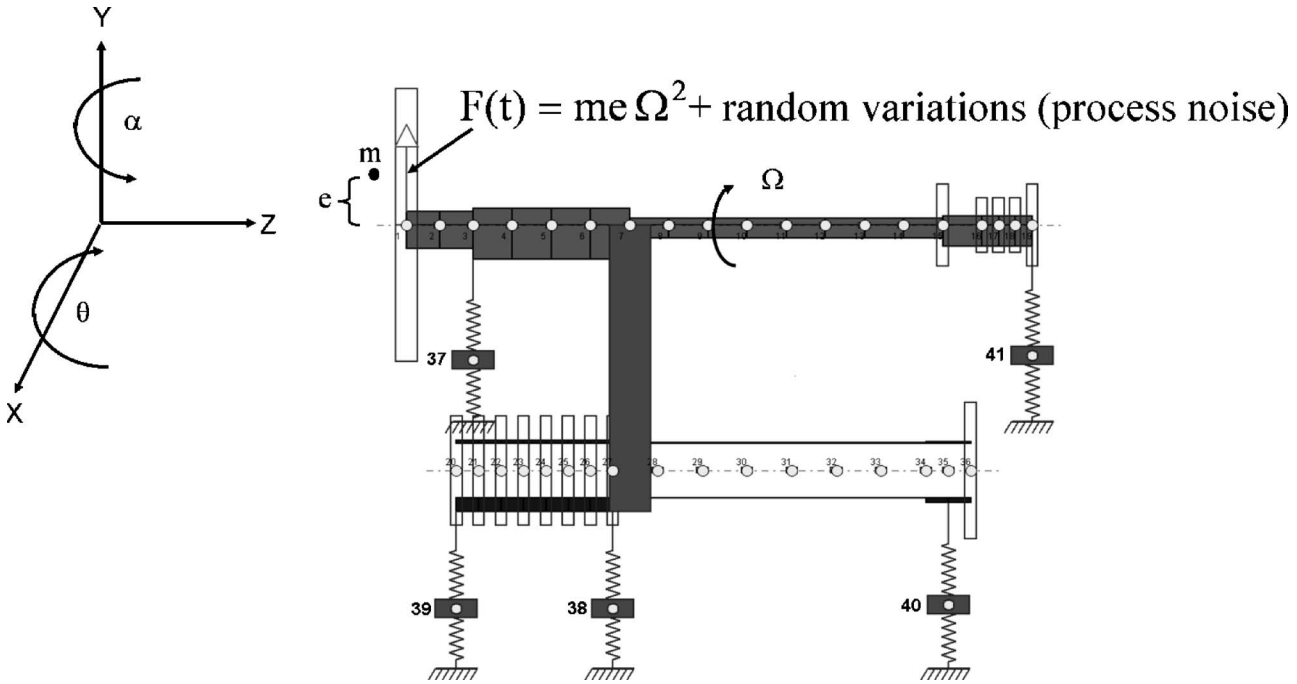


Fig. 3 Coordinate system used in the FEM and reduced order state space structural models. Cyclic imbalance forces are perpendicular to the *Z* axis. Impulse moments due to a FOD event are about the *X* axis.

fied rotor speed. Modal damping ratios are subsequently assigned to provide the best model fit to whirl-analysis frequency-response data calculated by DYROBES. An exact representation of the rotor response could be achieved by including both forward and backward modes of the rotor at the specified speed [11]. This technique, however, would not provide significantly better accuracy at a single rotor speed and would not lend itself to the simple state-space representation presented herein.

To obtain a reduced-order linear model with coefficients that are independent of Ω, i.e., with constant natural frequencies ω_{ni} $(i=1,\ldots,N)$ and modal damping, it is assumed that the natural frequencies and modes shapes calculated at the rotor critical speeds adequately account for the effects of gyroscopics over the operating range of interest. For the rotor described in Sec. 2, the ω_{ni} and φ_i are calculated using the DYROBES FEM code. The mode shapes φ_i obtained are orthogonal with respect to the mass and stiffness matrices and provide the following normalization:

$$\{\varphi_i\}^T[M]\{\varphi_i\} = 1 \tag{2.3}$$

$$\{\varphi_i\}^T[K]\{\varphi_i\} = \omega_{ni}^2 \tag{2.4}$$

Given the above-described assumptions, the equations of motion for the linearized rotational system take the form of Eq. (2.1). For lightly damped systems the matrix B is typically assumed to be diagonal [16]. With the eigenvectors mass-normalized and $[\Phi]=[\varphi_1\cdots\varphi_N]$, the coordinate transformation

$$\{q(t)\} = [\Phi]\{\eta(t)\} \tag{2.5}$$

is inserted into Eq. (2.1) and premultiplied by Φ^T. The dynamics of the system in modal coordinates are

$$\{\ddot{\eta}\} + \mathrm{diag}[2\xi_i\omega_{ni}]\{\dot{\eta}\} + \mathrm{diag}[\omega_{ni}^2]\{\eta\} = \Phi^T\{F\} \tag{2.6}$$

with the corresponding state-space form

$$\begin{Bmatrix} \dot{\eta} \\ \ddot{\eta} \end{Bmatrix} = \begin{bmatrix} 0 & I \\ -\mathrm{diag}[\omega_{ni}^2] & -\mathrm{diag}[2\xi_i\omega_{ni}] \end{bmatrix} \begin{Bmatrix} \eta \\ \dot{\eta} \end{Bmatrix} + \Phi^T \begin{Bmatrix} 0 \\ F \end{Bmatrix}$$

$$\{q\} = [\Phi \quad 0] \begin{Bmatrix} \eta \\ \dot{\eta} \end{Bmatrix} \tag{2.7}$$

and the modal damping ratios ξ_i defined as [16]

$$\xi_i = \frac{C_i}{C_{i\mathrm{critical}}} \tag{2.8}$$

For the present study, the coordinates of interest (Fig. 3) are

$$\{q\} = \begin{Bmatrix} y \\ \theta \end{Bmatrix} \tag{2.9}$$

at specified points (i.e., node locations in Fig. 1) in the rotor-bearing system.

The above presents a simplified formulation for the equations of motion of a rotational system which, in a linearized sense, include the gyroscopic effects. These effects are accounted for in DYROBES' analyses. Modal damping ratio estimates are based on whirl analysis log-decrement calculations at the rotor-bearing system critical speeds and the nominal speed of rotation. Comparison of the reduced-order structural model frequency response (to an assumed imbalance) and transient response to the FEM-generated data will provide a measure of the fidelity of the reduced-order model to appropriately describe a FOD-event induced structural response.

2.2 Foreign Object Damage Events Modeled Using Reduced Order Rotor Models From FEM Output Data. The state-space model presented in Eqs. (2.7)–(2.9) requires the modal or natural frequencies of the system and the eigenvectors (i.e., modal displacements) at specified locations on the rotor or bearings/supports. For the purposes of FOD detection algorithm development, the locations of interest would be those corresponding to possible vibration sensor locations, i.e., nodes corresponding to the bearings and supports as shown in Fig. 1. Commercial FEM codes can readily provide this information [13]. Transformation from modal coordinates to physical coordinates is performed via Eq. (2.5), with the matrix Φ formed by columns corresponding to the eigenvectors for each frequency being considered. To transform external forces on selected locations on the shaft (or supports) from physical to modal coordinates, the external force vector is premultiplied by Φ^T.

Physical (continuous) systems are inherently infinite order. Thus FEM, being lower order than the physical system being modeled, are implicitly inaccurate with FEM being more accurate in the low frequency range. The frequencies of interest are from 0 to 20,000 rpm, with the nominal rotational speed of the low pressure shaft being 9000 rpm. This frequency range will provide the required response due to fan imbalance, as well as the impulse response characteristics due to foreign object impact upon the fan disk. For adequate imbalance response over the full range of operating speeds, the natural frequencies and mode shapes obtained from the FEM correspond to the so-called critical speeds of the shaft [11]. To attain proper response to a FOD event, several of the lateral modes calculated via a FEM whirl analysis at a specified rotor speed are included in the model as well. The number of lateral modes included is determined by the performance of the ROM with respect to the FEM. Once again, the goal of the present study is to develop structural models to test real-time diagnostic algorithms, with these models providing acceptable response characteristics for the forcing functions of interest (imbalance and foreign object impact) while maintaining the model order low. The above-described reduced-order model development technique can be applied to FEM developed using codes intended for more elaborate analyses.

The reduced-order dynamic model was implemented in MATLAB®, using the companion software package SIMULINK®, in state-space form. Figure 4 depicts the SIMULINK structural model. Gas turbine rotors tend to have an inherently small degree of imbalance, which is modeled as an input corresponding to a variable amplitude sine wave with a frequency equal to the high- and/or low-speed rotor speeds. Imbalance force amplitude is frequency dependent according to the following relation:

$$F_{\mathrm{imbalance}} = me\Omega^2 \tag{2.10}$$

where m is the total mass of the rotating component, e is the effective eccentricity, and Ω is the shaft rotational speed in rad s^{-1}. If a FOD event results in additional imbalance, the eccentricity is increased by an additional amount. For the present study, the fan is assumed to have an eccentricity (i.e., inherent imbalance) of 0.001 in. (0.0025 cm).

A FOD event is modeled as an object hitting a point along the radius of the fan, resulting in a moment input at the node on the rotor corresponding to the fan location (Node 1 in Fig. 1). The magnitude of the input moment is determined as follows. Given a foreign object of mass m_{FO} with an initial linear velocity relative to the fan disk v_{FO} along the z axis (Fig. 3), conservation of angular momentum may be applied to determine an estimate of the initial angular velocity $\dot{\theta}$ of the fan disk about an axis perpendicular to the axis of rotation of the rotor (X axis in Fig. 3) due to the impact, i.e.,

$$\dot{\theta} = \frac{rm_{\mathrm{FO}}v_{\mathrm{FO}}}{I_\theta} \tag{2.11}$$

where I_θ is the moment of inertia of the fan disk about the X axis (Fig. 3) and r is the distance of the impact from the axis of rotation (Z axis in Fig. 3). The corresponding moment applied to the rotor at Node 1 from the impulse load on the fan disk is determined by the rate of change of angular momentum, given a specified time interval for the impact to occur Δt,

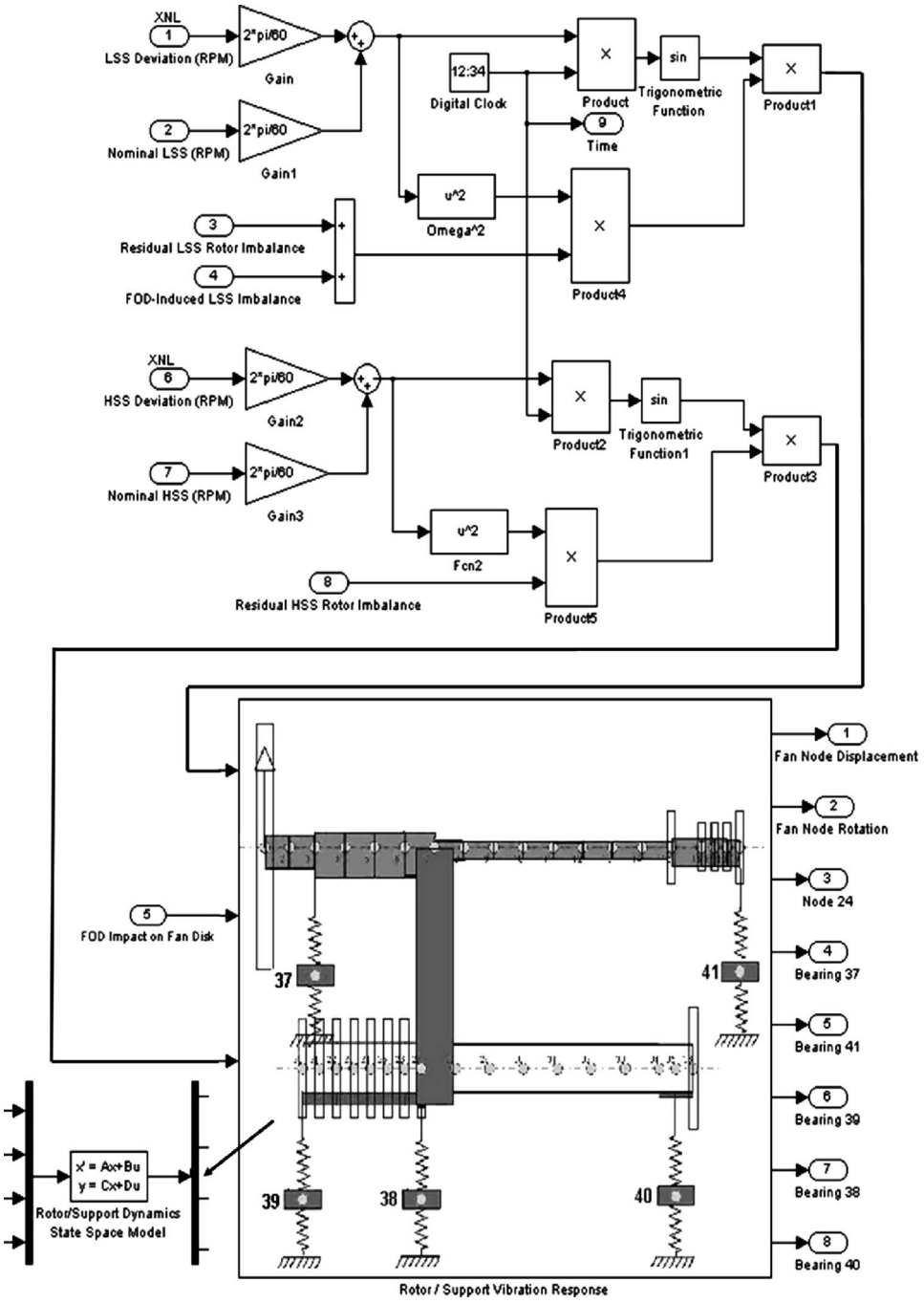


Fig. 4 MATLAB/SIMULINK **implementation of the reduced-order turbofan rotor/bearing system structural model**

$$T_\theta = I_\theta \frac{\Delta \dot{\theta}}{\Delta t} \tag{2.12}$$

The above-presented treatment provides a straightforward means for determining the impulse-load related torque input to the rotor during the initial stages of the FOD event, and is considered to be adequate within the level of complexity of the model needed for real-time diagnostic system development. As mentioned previously, information regarding specific FOD events does not appear in the open literature. The FOD event impact time interval, Δt, is dependent upon random factors such as foreign object geometry, location and angle of impact, foreign object hardness, relative speed, etc., and is estimated based on the authors' "intuitive feel" for an event that would cause damage.

Figure 5 presents a frequency response comparison of 4, 10, and 25 degree of freedom (DOF) reduced-order models with the frequency response of the 154 DOF DYROBES FEM (refer to Fig. 1 for node locations). Figures 6 and 7 present the displacement response of the 10 DOF ROM and FEM to a combined FOD event-imbalance input. Figure 6 presents the response at bearing/support location 37 to a "large" FOD event (m_{FO}=2 lbm (0.9 kg)) and a "small" FOD event (m_{FO}=0.25 lbm (0.11 kg)). Figure 7 presents the response at bearing/support location 39 to a "large" FOD event (m_{FO}=2 lbm (0.9 kg)) and a "small" FOD event (m_{FO}=0.25 lbm (0.11 kg)). The imbalance input is calculated using Eq. (2.10) with an eccentricity of 0.001 in. (0.0025 cm) con-

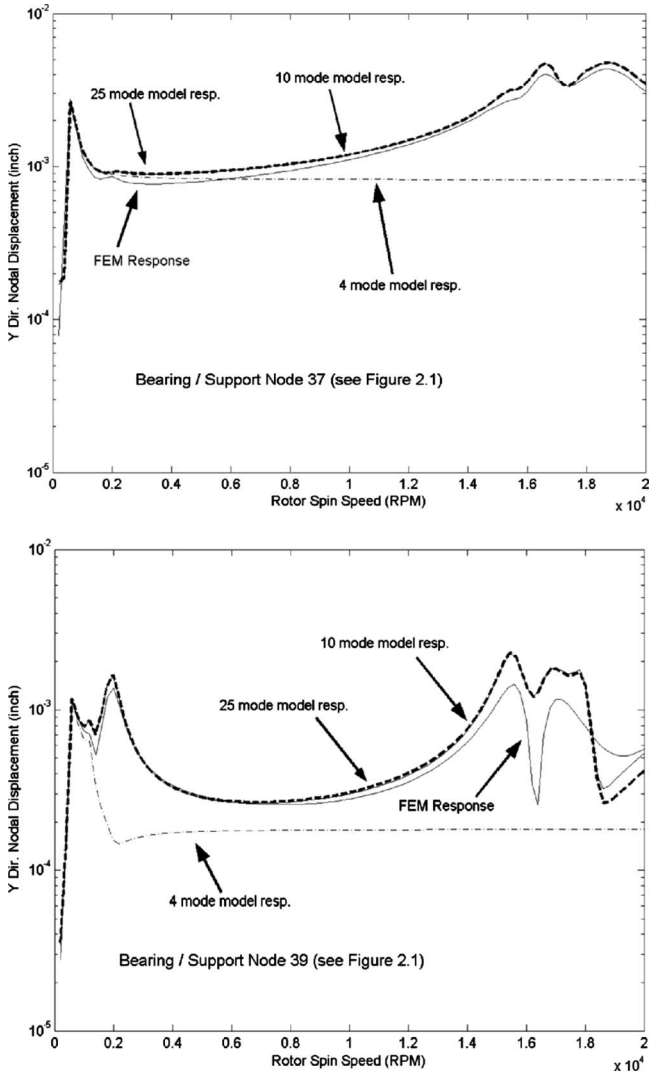


Fig. 5 Steady state imbalance response comparison of 4, 10, and 25 degree of freedom (DOF) reduced-order models for two different bearing/support locations. Imbalance has an eccentricity of 0.001 in. (0.0025 cm) located at the fan.

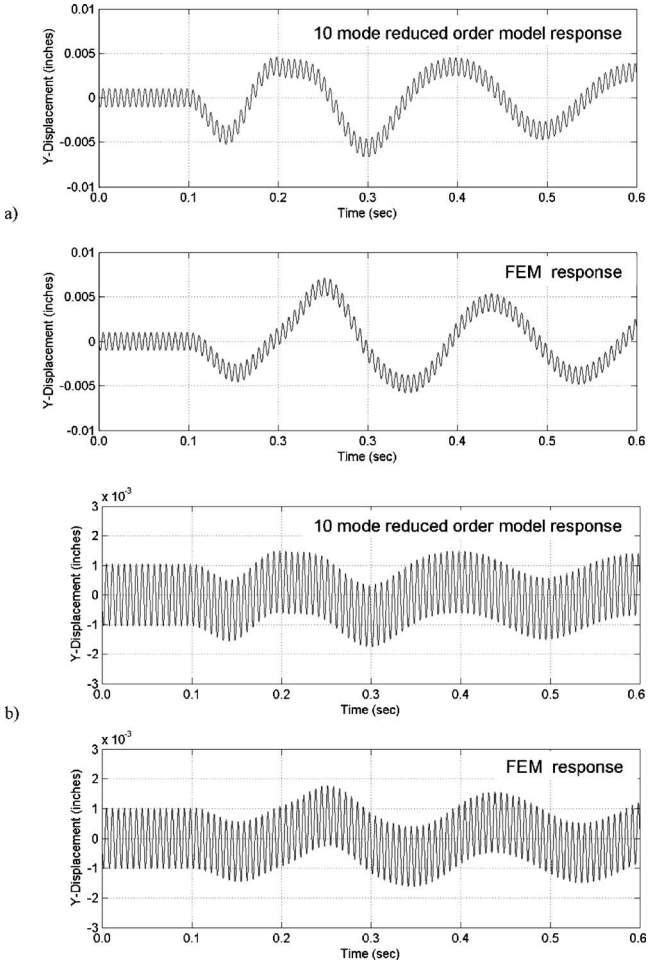


Fig. 6 Response at bearing/support location 37 to (*a*) a "large" FOD event (m_{FO}=2 lbm, (0.9 kg)) and (*b*) a "small" FOD event (m_{FO}=0.25 lbm (0.11 kg))

centrated at the fan node (node 1 in Fig. 1). The magnitude of the FOD event moment input at node 1 is calculated using Eq. (2.11) with a radius of 20 in. (51 cm), specified mass, a Δt of 0.1 s, and a relative velocity of 300 mph (134 m/s). To provide adequate FOD event response characteristics, the first lateral mode from a FEM whirl analysis at 9000 rpm was included. The 10 DOF model provides acceptable correspondence to the FEM and is considered sufficient for the present study. Differences between the reduced-order model and the FEM are due to the assumptions made, i.e., the ROM uses constant modal damping ratios and natural frequencies over the full range of operating speeds.

For both locations, the change in the vibration signal due to the large FOD event is easily discernible from the nominal vibration signal due to inherent imbalance alone. For the small FOD events, there is less of a difference from the normal operating characteristic. For smaller FOD events with the presence of sensor or process noise (which would be the case in an actual engine) the event could easily be masked. Thus, the signal analysis technique employed to detect such an event must be capable of identifying the characteristic vibration signature in the presence of process and sensor noise.

3 Optimal Sensor Placement for FOD Event Detection

Signal processing techniques that provide information concerning the internal modes (i.e., states) of a system require that these state variables be observable using the available sensor suite [17]. For FOD-event detection, the vibration sensor (e.g., accelerometer) signal must contain the modal response of the rotor in the frequency range of interest. The modes of a system are observable if and only if the observability matrix

$$O = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} \tag{3.1}$$

has rank n, where n is the dimension of the state space of the linear time invariant system under consideration, and C and A are its output and state matrices, respectively. For the 10 mode ROM representing the rotor-bearing system shown in Fig. 1 (with nodes 37 through 41 as candidate accelerometer locations) the system is completely observable. To determine the optimal (or worst case) accelerometer location for FOD event detection, a measure of the *degree of observability* of a system for a given sensor arrangement must be determined. One method suggested in [17] requires calculation of the Observability Grammian to determine which states, starting from a specified set of initial conditions, have an influence on the output energy $E(x(0))$ of the system over all time. Consider the following measure of the output energy at some

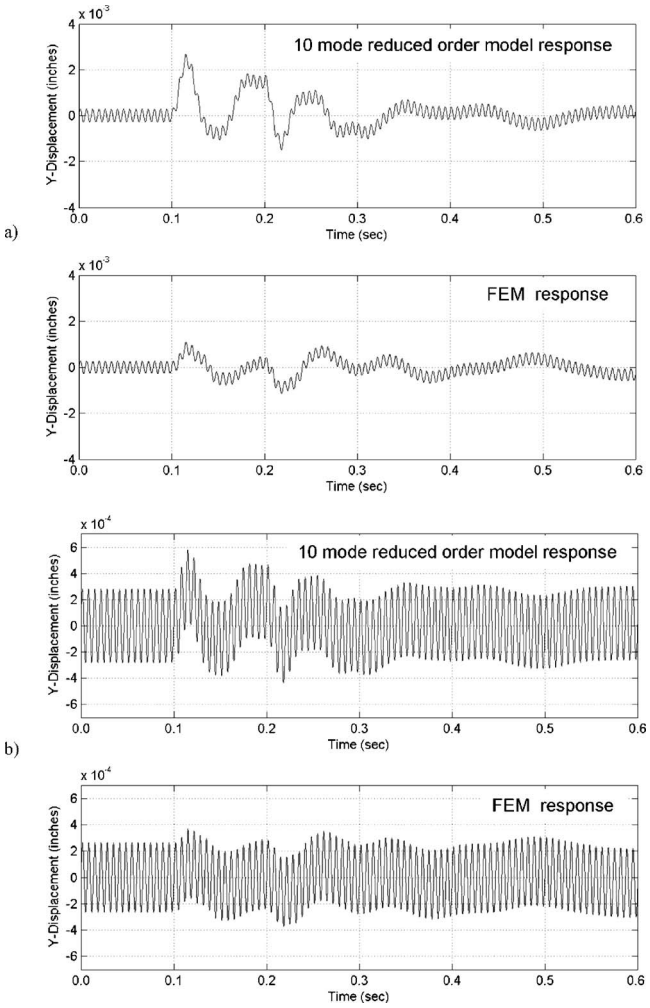


Fig. 7 Response at bearing/support location 39 to (*a*) a "large" FOD event (m_{FO}=2 lbm (0.9 kg)) and (*b*) a "small" FOD event (m_{FO}=0.25 lbm (0.11 kg))

initial condition $x(0)$ for the system of Eq. (2.3) under observation with no external input

$$E(x(0)) = \int_0^\infty \|y(t)\|^2 dt = \int_0^\infty y(t)^T y(t) dt = x(0)^T Q x(0) \quad (3.2)$$

where the Observability Grammian is defined as

$$Q = \int_0^\infty e^{A^T t} C^T C e^{At} dt \quad (3.3)$$

For a stable system, the existence of Eq. (3.3) is guaranteed and Q may be found by solving the following Lyapunov equation:

$$A^T Q + QA + C^T C = 0 \quad (3.4)$$

For a reduced number of accelerometers (reduced relative to the initial set, i.e., accelerometers 37 through 41 in Fig. 1) the system is observable if and only if Q is positive definite. With observability established for the reduced sensor suite system, Eq. (3.2) is used to determine the output energy contribution of each state variable. Large output energies correspond to easily observable state variables (modes). The optimal suite of sensors would naturally be that with the (relatively) largest output energies for the modes needed to identify a FOD event. Consider the calculation of Eq. (3.2) for locations 37 and 40. With the entry in the initial state vector (i.e., $x(0)$) for a specified mode set to 1.0 and the initial conditions for each of the other modes set to 0.0, $E(x(0))$ for each mode observed separately is simply the diagonal entry of Q corresponding to each mode. Thus, a judgment as to the optimal placement of the accelerometers may be made based on the diagonal entries of Q alone. For locations 37 and 40

$$\begin{aligned} \mathrm{diag}[Q]_{\text{node 37}} &= \begin{bmatrix} 4.1211\text{E}-003 \\ 5.1742\text{E}-003 \\ 2.3575\text{E}-005 \\ 9.3601\text{E}-007 \\ 4.4442\text{E}-006 \\ 3.9526\text{E}-004 \\ 2.5111\text{E}-003 \\ 3.6104\text{E}-004 \\ 5.0943\text{E}-003 \\ 9.8736\text{E}-004 \end{bmatrix} \\ \mathrm{diag}[Q]_{\text{node 40}} &= \begin{bmatrix} 3.6160\text{E}-007 \\ 3.1261\text{E}-005 \\ 1.4738\text{E}-005 \\ 6.9672\text{E}-003 \\ 4.4159\text{E}-004 \\ 4.3813\text{E}-004 \\ 1.6479\text{E}-004 \\ 2.9526\text{E}-004 \\ 2.2466\text{E}-008 \\ 1.0822\text{E}-004 \end{bmatrix} \end{aligned} \quad (3.5)$$

which, due to the higher observation energies associated with the first three modes, implies that placement of an accelerometer at location 37 may be more appropriate for detecting a FOD event. Upon performing the calculation of Eq. (3.2) for nodes 37 through 41 as candidate accelerometer locations, node 37 was determined to be the preferred location with node 40 being the least preferred. Thus, the analysis (i.e., feature extraction) technique selected should be applied to the signal generated by "accelerometer 37" in order to have the highest likelihood of detecting a FOD event. However, placement of a sensor at the optimal location may not be physically attainable, thus it is desirable that this same technique should be robust enough to detect the event at a suboptimal (possibly the most limiting) location. For some systems this conclusion may be drawn directly from physical intuition. For more complex systems (which may not easily lend themselves to physical intuition) the above-presented analytical treatment provides a quantitative measure for optimal sensor placement.

4 FOD Event Detection via Rotor Bearing Vibration Signal Analysis

Consider the simulated bearing accelerometer signals shown in Fig. 8. To demonstrate the robustness of the wavelet-based feature extraction technique, the analyses described in the following were performed on the signal from the most limiting accelerometer location on the rotor bearing system considered, which was determined from the sensor placement analysis of Sec. 3 to be at node 40 (Fig. 1). This location would result in the weakest signal and necessarily correspond to the most difficult location for FOD event detection. At 3.44 s, a 0.5 lbm (0.2 kg) foreign object hits the fan disk with a speed of 300 mph (134 m/s) (parallel to the Z axis in Fig. 3) at a radius of 20 in. (51 cm). The event is modeled as a pulse of width 0.04 s, with a magnitude determined using the method presented in Sec. 2.2. A fan disk eccentricity of 0.001 in. (0.0025 cm) is assumed. As shown in Fig. 8(*a*), the event is barely noticeable in the time trace for a noise-free situation. However

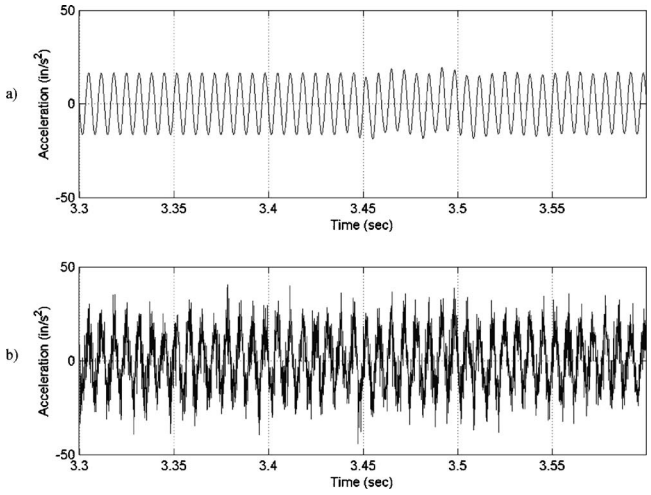


Fig. 8 Bearing 40 accelerometer output signal corresponding to a FOD event. (*a*) Noise free system response. (*b*) System response with process and sensor noise.

accelerometer signals on in-service engines are typically noisy. In addition to "process noise," i.e., random motion of the aircraft due to wind gusts and compensating maneuvers being transmitted from the airframe to the engine, there is a significant amount of sensor noise which may mask the occurrence of a FOD event. Figure 8(*b*) shows the accelerometer signal at location 40, with a signal-to-noise ratio of approximately 3.5 (11 dB). Observation of the signal shows no well-defined point in time at which one would identify a FOD event occurrence.

4.1 FOD Event Detection Using Fourier Analysis of Rotor Bearing Vibration Signals. Vibration signal analysis is an invaluable tool for identifying mechanical component problems. The most utilized tool in vibration analysis, i.e., the Fourier transform, essentially assumes that the signals being analyzed are of infinite duration [18,19]. When the continuous-time Fourier transform

$$Y(f) = \int_{-\infty}^{\infty} y(t)e^{-j2\pi ft}dt \tag{4.1}$$

or, more practically, its discrete-time counterpart

$$Y(f) = \sum_{n=-\infty}^{n=\infty} y(n)e^{-j2\pi fn} \tag{4.2}$$

is applied to a signal of finite duration, spectral leakage effects can significantly reduce the resolution of the corresponding power (frequency) spectrum given by

$$|Y(f)|^2 = \left| \sum_{n=-\infty}^{n=\infty} y(n)e^{-j2\pi fn} \right|^2 \tag{4.3}$$

Windowing of the original signal aids in mitigating the effects of finite signal length (spectral leakage), however if the component of the signal being analyzed has "compact support" (e.g., an impulse due to a FOD-event induced structural response), windowing of the signal may not significantly aid in identifying the occurrence of the event in time due to the Fourier transform basis functions, i.e., sinusoids, being themselves functions of infinite support and limited in their ability to decompose such a finite-duration signal. The short-time Fourier transform (and the corresponding time-dependent power spectrum) uses a windowed version of $y(n)$ to compute $Y(f)$ in an attempt to localize a change in a signal over time [19]. The time-dependent power spectrum for the noise-free signal (Fig. 9(*a*)) shows a significant change at approximately 3.3 s, which disappears as time progresses. The time-dependent power spectrum of the noisy signal (Fig. 9(*b*)) shows no noticeable change in time, with the effect of the FOD event "buried" in the spectra due to the noise. Thus, there is a need for an alternative signal analysis technique for FOD event detection.

4.2 FOD Event Detection Using Wavelet Analysis of Rotor Bearing Vibration Signals. Over the past decade the discrete time wavelet transform (DTWT) has been applied to a wide range of signal analysis problems, e.g., denoising of signals as well as time localization and reconstruction of short duration changes [8]. Because of these characteristics, the DTWT is considered to be a viable candidate for identifying a short duration change in a bearing vibration signal due to a FOD event corrupted by noise.

The DTWT is the discrete-time counterpart of the continuous-time wavelet transform

$$CWT(a,\tau) = \frac{1}{\sqrt{a}} \int y(t)\Psi\left(\frac{t-\tau}{a}\right)dt \tag{4.4}$$

where $\Psi(t)$ is the basic or mother wavelet and $\Psi((t-\tau)/a)/\sqrt{a}$ are the wavelet basis functions which are scaled via the parameter a (by compressing or stretching the mother wavelet) and shifted in time by τ. The operation performed in Eq. (4.4) on the original signal $y(t)$ may be interpreted as the cross correlation of the signal $y(t)$ with $\Psi(t/a)/\sqrt{a}$, shifted by τ/a [18]. Thus, the CWT computes the component of $y(t)$ that is similar to $\Psi(t/a)/\sqrt{a}$. If little (or no) similarity exists between the two, then the CWT will be small (or zero). Larger values of CWT indicate better correlation. Figure 10 presents an overview of the wavelet transform process with multiple scale decomposition. The wavelet family used is the Daubechies family.

Typically, the resulting data are presented in terms of time (or k if in discrete time) and scale (i.e., coefficient) of the chosen wavelet transform. "Pseudo" frequency components of $y(t)$, corresponding to the scale of the constituent wavelets, F_m, are obtained by the following expression: [20]

$$F_m = \frac{TF_c}{m} \tag{4.5}$$

where m is the scale, T is the sampling interval, and F_c is the wavelet center frequency calculated by taking the maximum of the modulus of the wavelet's Fourier transform.

Most of the functions for $\Psi(t)$ have been developed over the past two decades with the Haar wavelet being the first documented function to be used in this context [19]. Choice of $\Psi(t)$ is heavily dependent on the nature of the signal (or a localized feature of a signal) being analyzed. For example, if a vibration signal were to be analyzed to determine whether a FOD event had occurred, one would expect to see an impulse-like localized feature embedded in a normally noisy sinusoidal signal. Intuitively, one would choose a wavelet that had an impulse-like character as opposed to one that resembled a step function, as is the case with the Haar wavelet mentioned previously.

In discrete-time the wavelet transform takes the form

$$\mathrm{DTWTD}(m,n) = a_0^{-\frac{m}{2}} \sum_k y(k)\psi(a_0^{-m}k - n\tau_0) \tag{4.6}$$

where m is the scale (degree of dilation) and n corresponds to the net translation in time (in terms of the number of time τ_0) of the wavelet at a specific scale. For scale $m=0$, the coefficients of the so-called scaling function, from which all of the wavelets in a given family are related, are

$$\mathrm{DTWTA}(0,n) = \sum_k y(k)\varphi(k - n\tau_0) \tag{4.7}$$

$\varphi(k-n\tau_0)$ is typically determined via iteration using the following dilation equation:

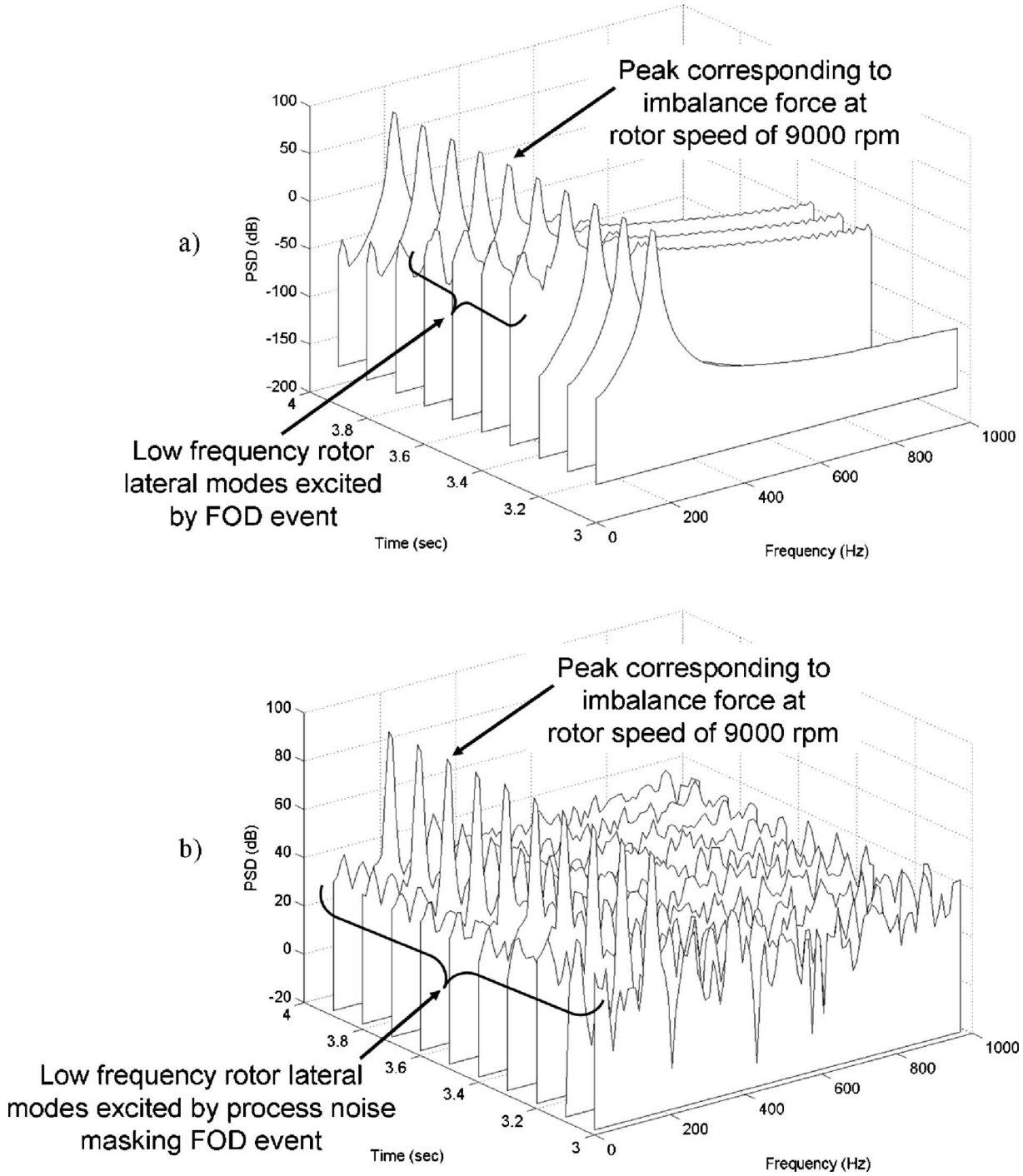


Fig. 9 Power spectra of bearing 40 accelerometer output signal corresponding to a FOD event. (*a*) Noise free system response. (*b*) System response with process and sensor noise.

$$\varphi(k) = \sum_{n=0}^{N-1} c_n \varphi(2k - n\tau_0) \tag{4.8}$$

where N is the number of terms in $\varphi(k)$. The scaling function $\varphi(k)$ determines the character of the wavelet. The characteristic of interest in the signal being analyzed typically dictates the form of the wavelets used. A wavelet $\psi(k)$ is related to its associated $\varphi(k)$ via

$$\psi(k) = \sum_{n=0}^{N-1} (-1)^n c_n \varphi(2k + n\tau_0 - N + 1) \tag{4.9}$$

Equation (4.9) is also a dilation equation with $\psi(k)$ referred to as a *dilation wavelet* [19]. The scaling function, its dilates, the corresponding wavelets and their dilates are orthogonal to one another and therefore constitute a set of functions by which an arbitrary function may be built [19], completely analogous to the Fourier transform and series. The base value of scaling parameter, a_0, and the time shift, τ_0, are typically set equal to 2 and 1, respectively, for computational efficiency, analogous to the fast Fourier transform.

The convolution of Eqs. (4.6) and (4.7) is commutative, i.e.,

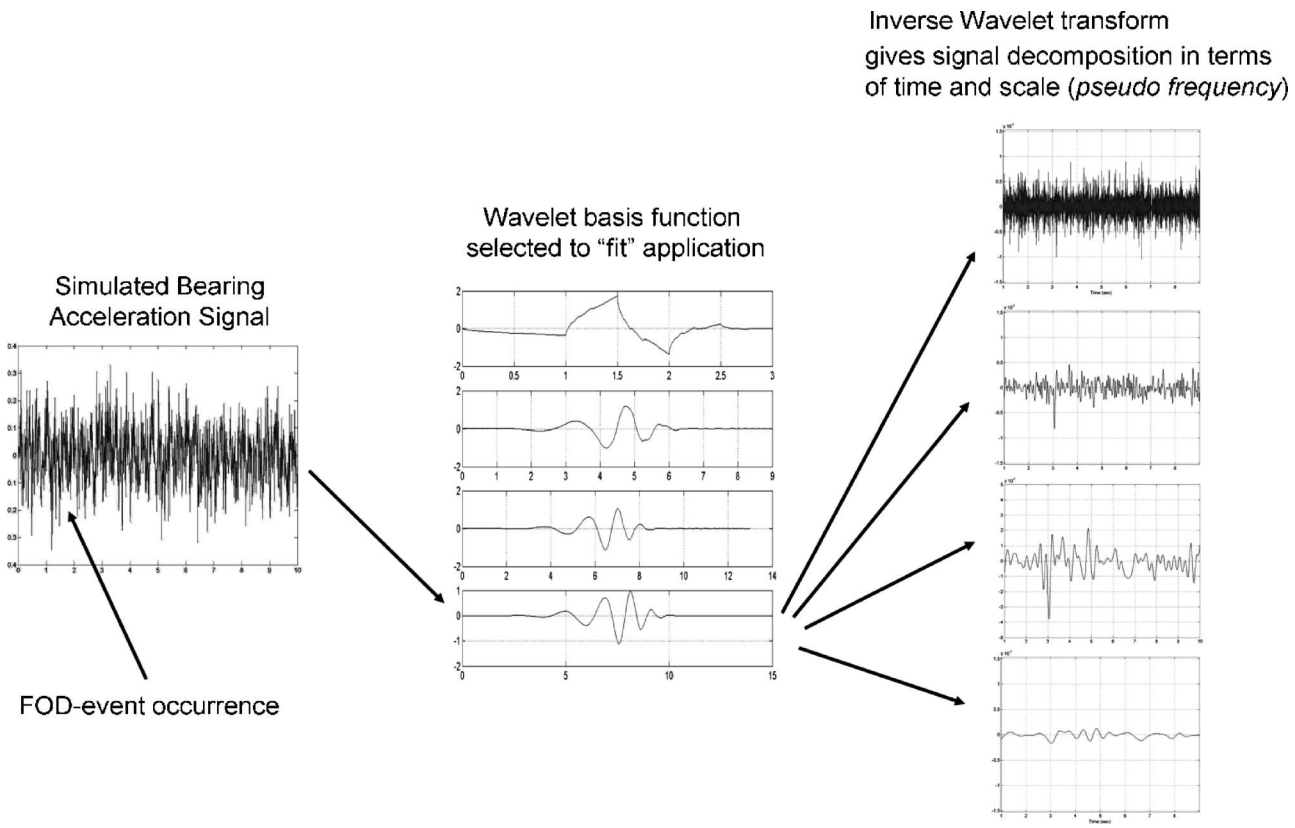


Fig. 10 Example bearing vibration signature and corresponding wavelet decomposition accompanying a FOD event

$$\mathrm{DTWTA}(0,n) = \sum_k \varphi(k) y(k - n\tau_0) \tag{4.10}$$

$$\mathrm{DTWTD}(m,n) = a_0^{-\frac{m}{2}} \sum_k \psi(k) y(a_0^{-m}k - n\tau_0) \tag{4.11}$$

Thus, the discrete-time wavelet transform may be viewed as a filtering operation with $\varphi(k)$ containing the coefficients of a low-pass finite impulse response (FIR) filter and $\psi(k)$ the coefficients of a high-pass FIR filter. The coefficients corresponding to the low-frequency portion of the input signal, i.e., *approximation coefficients* (Eq. (4.10)), and the high-frequency *detail coefficients* (Eq. (4.11)) at a given *scale* in a wavelet decomposition are provided by successive low-pass and high-pass filtering operations, respectively. Reconstruction of the approximations and details at a specified scale m is performed via the corresponding inverse wavelet transform

$$D_m(k) = \sum_n \mathrm{DTWTD}(m,n)\psi_{m,n}(k) \tag{4.12}$$

$$A_m(k) = \sum_n \mathrm{DTWTA}(m,n)\varphi_{m,n}(k) \tag{4.13}$$

and is implemented via banks of *reconstruction* or *synthesis* filters. Figure 11 presents a two-scale DTWT-based signal analysis and synthesis using a FIR quadrature mirror filter bank. Locations in the filter bank correspond to the signal characteristics of interest (details or approximations). Scaling is accomplished by downsampling after high- or low-pass filtering. Successive upsampling and filtering performs the reconstruction. When the reconstruction filter coefficients correspond to the *quadrature mirror filter* of the decomposition filter, perfect signal reconstruction is guaranteed [21]. This is the practical implementation of the discrete-time wavelet transform algorithm developed by Mallat [22–24]. The information gained from the decomposition depends on the application e.g., a signal may be "denoised" by omitting the detail from one or more scales during the reconstruction. As well, abrupt changes in the signal, e.g., a FOD event-induced accelerometer response, might be detected by observing one or more of the approximations.

Figure 12 presents a wavelet analysis of the signal analyzed in Sec. 4.1, the noise-free bearing accelerometer signal at location 40. The Daubechies 8 wavelet was chosen (a subjective choice) for the analysis due to its correspondence to the signal characteristic of interest. Observing the approximation at scale 8 (Fig. 12(*b*)) demonstrates the wavelet's ability to determine the location in time of the event. Several other approximations and details at various scales (not shown) also revealed the occurrence of a FOD

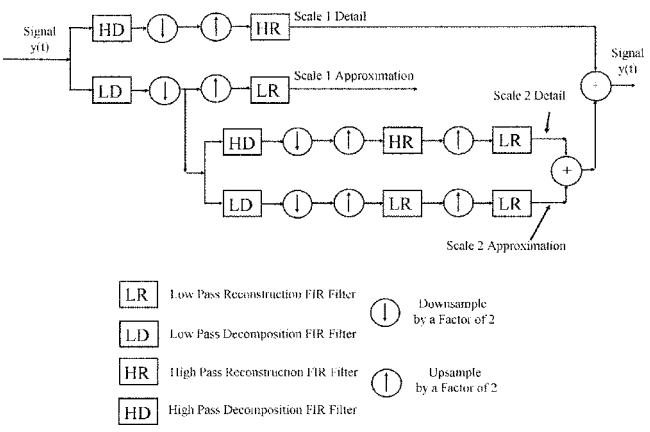


Fig. 11 Example of a two-scale DTWT-based signal analysis and synthesis using a FIR quadrature mirror filter bank

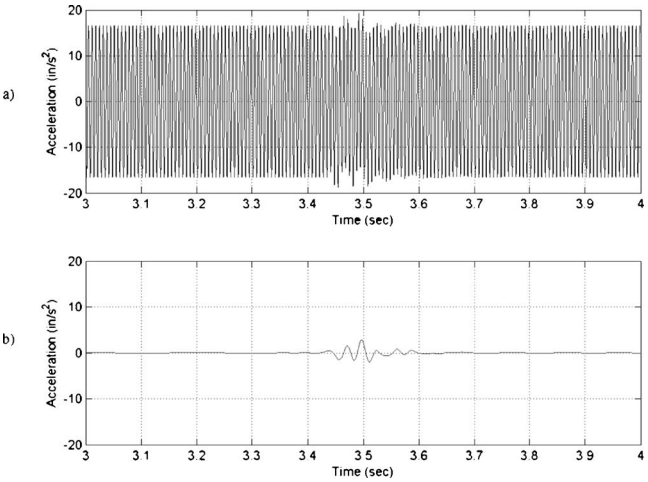


Fig. 12 Wavelet analysis of noise-free bearing accelerometer signal at location 40. (*a*) Accelerometer signal. (*b*) Corresponding Daubechies 8 wavelet inverse transform, scale 8 approximation.

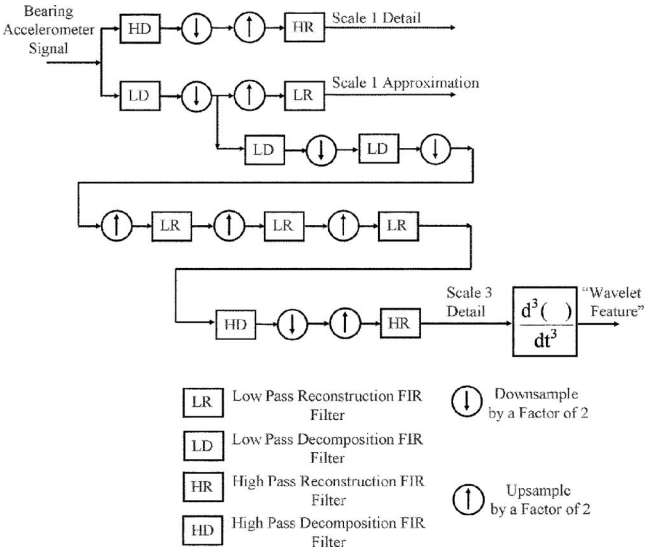


Fig. 14 Wavelet transform-based vibration feature extraction implemented using an analysis and synthesis FIR filter bank

event. The analysis of the noisy signal at the same scale's approximation (Fig. 13), however, illustrates that the wavelet chosen (as well as several other wavelets tested, not presented here), may have difficulty highlighting a subtle change buried in a low signal-to-low noise ratio (SNR=3.5) signal. Additional "conditioning" of the decomposed signal is required.

Abrupt changes in an otherwise smooth signal will correspond to a localized change in the signal derivative with respect to time (or some space-related dependent variable depending on the application). The more abrupt the change, the larger the local value of the derivative with respect to time. Additionally, for the same change the value of higher order derivatives

$$\frac{d^n y(t)}{dt^n} \tag{4.14}$$

will provide a frequency-dependent gain, i.e., the more abrupt the signal change, the larger the gain. This quality of the derivative is more easily observed in the frequency domain by taking the Laplace transform of Eq. (4.14) and setting $s=j\omega$,

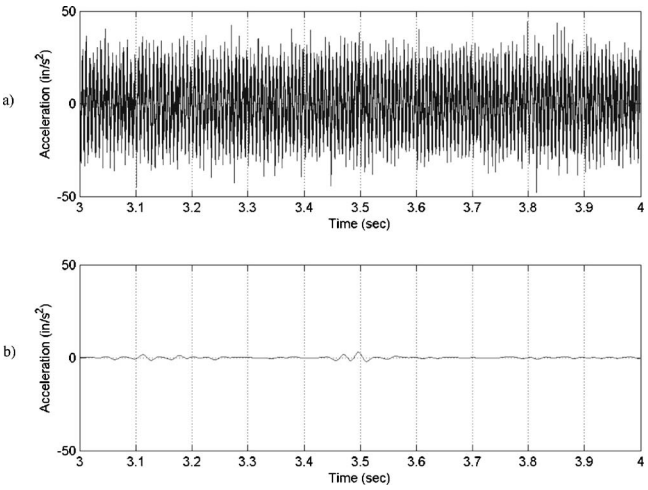


Fig. 13 Wavelet analysis of noisy bearing accelerometer signal at location 40. (*a*) Accelerometer signal. (*b*) Corresponding Daubechies 8 wavelet inverse transform, scale 8 approximation.

$$L\left\{\frac{d^n y(t)}{dt^n}\right\}_{s=j\omega} = (j\omega)^n y(j\omega) \tag{4.15}$$

The corresponding frequency-dependent magnitude is

$$\mathrm{MAG}((j\omega)^n y(j\omega)) = \omega^n \mathrm{MAG}(y(j\omega)) \tag{4.16}$$

Thus, multiple differentiations in the time domain provides an "s^n" multiplication factor in the frequency domain. The higher frequency characteristics of a signal (e.g., a FOD event would be classified as a high-frequency event) would have a significantly higher gain relative to the low-frequency components. It would seem that in a noise-free signal, the occurrence of a FOD event resulting in an impulsive input to the turbofan rotor of the present study could easily be identified using this technique. Multiple differentiation of noisy signals is typically avoided in most signal processing applications due to the amplification provided by Eq. (4.16). Indeed, multiple differentiations of the signal shown in Fig. 13(*a*) would make identification of a FOD event virtually impossible, primarily due to amplification of the noise.

Mallat [25] proposed a technique for identifying a wavelet specifically designed for edge detection in computer vision systems. A signal (or 2D image in the original application) is passed through a smoothing function (e.g., a low-pass filter) and differentiated multiple times. The combination of smoothing and differentiation results in a wavelet customized to the application. For this research, the technique was adapted to identify the FOD event induced short-time change in accelerometer signals. The smoothing filter chosen for the present investigation is the Daubechies-8 wavelet decomposition. As shown in Fig. 14, the accelerometer signal is passed through an approximation filter bank three times and subsequently passed through a detail filter. This provides a means to focus on a *subband* of the original signal where the feature of interest is thought to lie, the net effect being a low-pass filtering or smoothing operation. The result is differentiated three times to produce the wavelet-based feature of interest. Figure 15 presents the wavelet feature extracted from the accelerometer signal shown in Fig. 13, using the above-described custom wavelet. The feature extracted is the spike shown in Fig. 13(*b*). The exact time of the event is precisely determined using this technique. Indeed, the event is modeled by imposing a pulse, whose magnitude is a function of the foreign object impact characteristics mentioned previously, on the fan disk. The two spikes shown in Fig. 15 result from the rising and descending parts of that pulse "filtered" through the mechanical (i.e., rotor-bearing) system. The

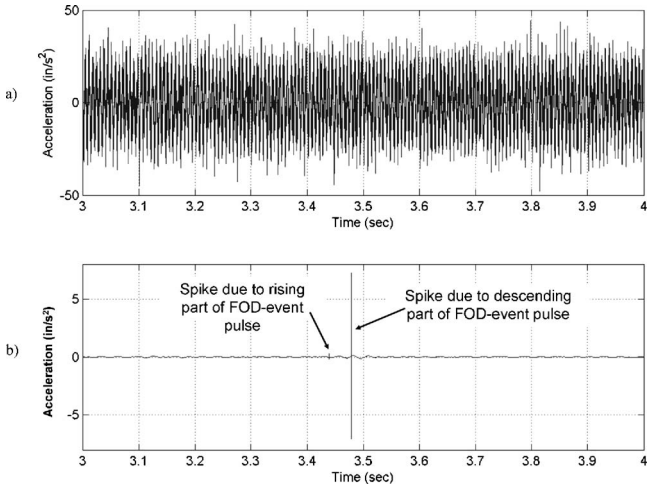


Fig. 15 Wavelet feature extracted from noisy bearing accelerometer signal at location 40. (*a*) Accelerometer signal. (*b*) Corresponding conditioned Daubechies 8 wavelet inverse transform, scale 8 approximation.

event input profile to the model (i.e., the above-described pulse input) was considered to be appropriate for a significant, *structurally damaging* event, i.e., where the foreign object absorbs virtually none of the energy of impact, resulting in complete and immediate transfer of energy to the fan disk. This would, for example, be the case for hard objects such as ice. For other scenarios, pulses of lower magnitude or with noninfinite rising (or descending) slopes may require higher-order differentiation for detection.

In summary, the result of the following sequence of operations:

(1) wavelet decomposition of a noisy (e.g., bearing accelerometer) signal,
(2) reconstruction of selected components (approximations *or* details at a selected scales) of the signal, and
(3) multiple differentiation of the reconstructed components

can precisely time-locate an abrupt, high-frequency, change in the signal. However, choice of too high a scale may result in the high-frequency event being filtered out of the signal.

5 Conclusion

This paper describes the development of a wavelet-based feature specifically designed to identify the occurrence of FOD events in gas turbine engines. The wavelet-based feature is extracted from rotor bearing accelerometer signals and is shown to be robust in the presence of significant noise. The technique is intended ultimately to be combined with analytical measurements (e.g., Kalman filter thermal/health parameter estimates) for FOD-event detection via information fusion from these (and perhaps other) sources. It is envisioned that fusion of structural and thermal performance information will provide a pilot, or an autonomous vehicle, more confidence in a FOD-event diagnosis, thus enabling a corrective action to be based on a more informed decision.

Due to the lack of high-frequency FOD-event test data in the open literature, a reduced-order turbofan structural model (ROM) was synthesized from a finite-element model to support the investigation. The wavelet-based FOD-event detection scheme developed is "data driven" (i.e., not model-based), thus precise correspondence of the ROM to the FEM (which itself would probably not represent an actual engine with complete fidelity) is not essential. The ROM is considered to be "phenomenologically accurate," and shown to provide acceptable fidelity with regards to the application, i.e., real-time diagnostic system development.

Using the state-space ROM, a sensor placement analysis was performed to determine the optimal accelerometer location for FOD-event detection. To demonstrate the robustness of the technique, testing was performed using the least optimal bearing location. For comparison, analysis of noise-free and noisy accelerometer signals was performed using a standard wavelet decomposition and shown to be inadequate for detecting FOD events. In the presence of significant noise (SNR=3.5, 11 dB), precise location of the FOD event in time was obtained using the custom wavelet feature developed. Choice of the smoothing wavelet used (Daubechies 8) was subjective, and based on knowledge of the signal characteristic of interest, i.e., a FOD event-induced abrupt change in an otherwise noisy signal.

Future work should include determining the optimal wavelet (and corresponding scale) to apply based on current information theoretic techniques such as maximum entropy. Additionally, combination of structural and thermal performance information could provide a robust diagnosis of foreign object damage in specific engine components, i.e., extending this research to the fault detection isolation problem. Continued development and verification of the feature presented would require high-frequency accelerometer data (i.e., sensor bandwidth greater than 10 kHz) from FOD event engine tests at multiple operating conditions.

Nomenclature

The nomenclature used in this paper is standard for the specific disciplines described and thus symbols may have multiple meanings. The correct meaning should be clear from the context, but this section relates symbol definitions to the corresponding sections of the paper.

B = damping coefficient matrix
C_i, C_{critical} = damping
CDS = control and diagnostic system
e = effective eccentricity
F = force
G = gyroscopic matrix
I_θ = moment of inertia about the X axis
K = stiffness, stiffness matrix
M = mass matrix
m = mass of rotating component
m_{FO} = mass of foreign object
N = number of modes included in the model
q = displacement in physical coordinates
r = distance of impact from axis of rotation
v_{FO} = velocity of foreign object
η = displacement in modal coordinates
θ = rotation about X axis
ξ = modal damping ratio
T = moment
Φ = matrix composed of the ordered vectors φ
φ = mode shape (eigenvector) corresponding to ω_n
Ω = shaft rotational speed
ω_n = natural frequency (eigenvalue)
A = state matrix
C = output matrix
$E(\cdot)$ = output energy
n = order of the linear time invariant system
O = observability matrix
Q = Observability Grammian
A_m = reconstructed approximation component of wavelet-analyzed signal at scale m
a_0 = wavelet base value scaling parameter
c_n = scaling function coefficient
CWT = continuous-time wavelet transform
D_m = reconstructed detail component of wavelet-analyzed signal at scale m
DTWT = discrete-time wavelet transform

DTWTD = discrete-time wavelet transform detail coefficient
DTWTA = discrete-time wavelet transform approximation coefficient
f = frequency (Hz)
F_C = wavelet center frequency
F_m = constituent wavelets
m = dyadic scale of wavelet-decomposed signal
n = wavelet time delay in number of samples
N = number of terms in φ
T = sampling interval
$Y(f)$ = discrete Fourier transform of $y(n)$
τ = wavelet time shift
$\varphi_{m,n}$ = wavelet scaling function at scale m and translation n
$\psi_{m,n}$ = wavelet at scale m and translation n
ω = angular frequency (rad/s)

References

[1] Society of Automotive Engineers, 1998, "Uncontained Turbine Engine Rotor Events Data Period 1984 Through 1989," SAE SP-1270.
[2] Mattingly, J. D., Heiser, W. H., and Pratt, D. T., 2002, *Aircraft Engine Design*, 2nd ed., American Institute of Aeronautics and Astronautics, Reston, VA.
[3] Kerr, L. J., Nemec, T. S., and Gallops, G. W., 1991, "Real-Time Estimation of Gas Turbine Engine Damage Using a Control Based Kalman Filter Algorithm," ASME 91-GT-216, International Gas Turbine and Aeroengine Congress and Exposition, Orlando, FL, June 3–6.
[4] Luppold, R. H., Roman, J. R., Gallops, G. W., and Kerr, L. J., 1989, "Estimating In-Flight Performance Variations Using Kalman Filter Concepts," AIAA-89-2584, AIAA 25th Joint Propulsion Conference, Monterey, CA, July 10–12.
[5] Klein, L. A., 1999, *Sensor and Data Fusion Concepts and Applications*, 2nd ed., SPIE Press, Bellingham, WA, Vol. TT35.
[6] Doebelin, E. O., 2004, *Measurement Systems Application and Design*, 5th ed., McGraw-Hill, New York.
[7] Lim, M. H., and Salman Leong, M., 2003, "Diagnosis for Loose Blades in Gas Turbines Using Wavelet Analysis," *Proceedings of the ASME Turbo Expo 2003*, Atlanta, GA, June 16–19.
[8] Juluri, N., and Swarnamani, S., 2003, "Improved Accuracy of Fault Diagnosis of Rotating Machinery Using Wavelet De-Noising and Feature Selection," *Proceedings of the ASME Turbo Expo 2003*, Atlanta, GA, June 16–19.
[9] Bay, J. S., 1999, *Fundamentals of Linear State Space Systems*, McGraw-Hill, New York.
[10] The Math Works Inc., 1996, *MATLAB Control System Toolbox Users Guide*, Natick, MA.
[11] Ehrich, F. F., 1999, *Handbook of Rotordynamics*, Krieger, Malabar, FL.
[12] Botman, M., 1980, "Vibration Aspects of Small Turbine Engines in Aircraft Installations," Society of Automotive Engineers Turbine Powered Executive Aircraft Meeting, Phoenix, AZ, Apr. 9–11.
[13] Gunter, E. J., and Chen, W. J., 2003, *DyRoBeS Dynamics of Rotor-Bearing Systems, Windows Version 8, Users' Manual*, RODYN Vibration Analysis, Inc., Charolottesville, VA.
[14] Ogata, 1990, *Modern Control Engineering*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ.
[15] Lawrence, C., Kelly, C., and Gallardo, V., 2001, "Simulation of Aircraft Engine Blade-Out Structural Dynamics," NASA/TM-20010210957/REV1, NASA-Glenn Research Center, Cleveland, OH.
[16] Hatch, M. R., 2000, *Vibration Simulation Using MATLAB and ANSYS*, Chapman and Hall/CRC, New York.
[17] Bishop, R. H., 2002, *The Mechatronics Handbook*, CRC, New York.
[18] Chan, Y. T., 1995, *Wavelet Basics*, Kluwer Academic, Boston, MA.
[19] Newland, D. E., 1993, *An Introduction to Random Vibrations, Spectral and Wavelet Analysis*, 3rd ed., Wiley, New York.
[20] The MathWorks Inc., 2000, *MATLAB Wavelet Toolbox Users Guide*, Natick, MA.
[21] Proakis, J. G., Rader, C. M., Ling, F., and Nikias, C. L., 1992, *Advanced Digital Signal Processing*, Macmillan, New York.
[22] Mallat, S. G., 1989, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," IEEE Trans. Pattern Anal. Mach. Intell., **11**(7), pp. 674–693.
[23] Strang, G., 1989, "Wavelets and Dilation Equations: A Brief Introduction," SIAM Rev., **31**(4), pp. 614–627.
[24] Soman, A. K., and Vaidyanathan, P. P., 1993, "On Orthonormal Wavelets and Paraunitary Filter Banks," IEEE Trans. Signal Process., **41**(3), pp. 1170–1183.
[25] Mallat, S., and Hwang, W. L., 1992, "Singularity Detection and Processing With Wavelets," IEEE Trans. Inf. Theory, **38**(2), pp. 617–643.

Convergent Zone-Refinement Method for Risk Assessment of Gas Turbine Disks Subject to Low-Frequency Metallurgical Defects

Harry R. Millwater
Department of Mechanical Engineering,
The University of Texas at San Antonio,
One UTSA Circle,
San Antonio, TX, 78249
e-mail: harry.millwater@utsa.edu

Michael P. Enright
Reliability and Material Integrity,
Southwest Research Institute,
6220 Culebra Rd.,
San Antonio, TX, 78228
e-mail: menright@swri.org

Simeon H. K. Fitch
Mustard Seed Software,
1634 Brandywine Drive,
Charlottesville, VA, 22901
e-mail: simeon.fitch@mseedsoft.com

Titanium gas turbine disks are subject to a rare but not insignificant probability of fracture due to metallurgical defects, particularly hard α. A probabilistic methodology has been developed and implemented in concordance with the Federal Aviation Administration (FAA) Advisory Circular 33.14-1 to compute the probability of fracture of gas turbine titanium disks subject to low-frequency metallurgical (hard α) defects. This methodology is further developed here to ensure that a robust, converged, accurate calculation of the probability is computed that is independent of discretization issues. A zone-based material discretization methodology is implemented, then refined locally through further discretization using risk contribution factors as a metric. The technical approach is akin to "h" refinement in finite element analysis; that is, a local metric is used to indicate regions requiring further refinement, and subsequent refinement yields a more accurate solution. Supporting technology improvements are also discussed, including localized finite element refinement and onion skinning for zone subdivision resolution, and a restart database and parallel processing for computational efficiency. A numerical example is presented for demonstration. [DOI: 10.1115/1.2431393]

Introduction

As a result of the airliner accident at Sioux City, Iowa, in 1989 [1], where a disk rupture in the tail engine caused the loss of hydraulics and subsequent crash of the plane, the FAA requested in 1991 that industry, through the Aerospace Industries Association, review available techniques to determine whether a probabilistic damage tolerance approach could be introduced to reduce the rate of uncontained rotor events. In this case, the titanium fan disk fractured due to a crack that initiated from a hard α defect formed during the titanium melt and not detected during in-service inspections. The industry working group concluded that additional enhancements to the conventional rotor life management methodology could be established which explicitly address anomalous conditions.

The Federal Aviation Adminstration (FAA), with industry input, has developed an advisory circular (AC 33.14-1) [2] that, among other things, outlines a probabilistic damage tolerance analysis methodology for assessing the risk of fracture of titanium rotors on gas turbine engines. The probabilistic methodology complements and does not replace preexisting approaches.

The advisory circular (AC) advises that all new titanium rotors be designed such that the probability of fracture (POF) per flight is less than the design target risk (DTR). The DTR is the standard against which probabilistic risk assessment results (stated in terms of component event rates and engine level event rates) are compared. The DTR is currently set at 10^{-9} for the component event rate and 5×10^{-9} for the engine event rate, i.e., the average probability of fracture per flight over the life of the titanium component should be $<10^{-9}$ and the cumulative average probability of fracture per flight over the life of all engine titanium components should be less than 5×10^{-9}. A new engine design is expected to meet the DTR before the engine is certified.

The probabilistic damage tolerance requirement for hard α defects is a recent development and, as such, methodologies to quantify the POF are also fairly recent. A methodology for high-frequency anomalies in nickel-based powder metal components has been developed and validated by comparison to seeded fracture specimens [3]. Although developed for powder nickel, the probabilistic equations will reduce to the simpler form described in the AC and implemented here; however, the computational strategy for solving the equations is different.

The advisory circular specifies one approach to quantifying the POF. This methodology is based on discretizing the disk structure into a number of zones. A zone is a grouping of material such that all locations in the zone have a generally uniform stress state, and the same fatigue crack growth properties, inspection schedules, probability of detection curves, and anomaly distribution. In other words, the risk computed for any subregion of material of the zone will be the same or nearly so, and therefore, the risk is assumed constant for a defect located anywhere in that zone. Since there will be some variation in calculated POF, depending on the location of the defect within a zone, a conservative approach is specified whereby the defect is placed in the life-limiting location, i.e., the location within the zone that will yield the highest POF. The finite element mesh and stress results are used as the framework to select zone boundaries and, hence, the zone discretization.

The FAA has funded a research program in consort with several commercial engine original equipment manufacturers and research organizations to develop computational techniques and software to quantify the POF due to hard α defects in titanium based on the methodology outlined in the AC. The probabilistic methodology from this research program is outlined below and discussed in detail in [4,5]. Subsequent work is ongoing to extend

Submitted to ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received March 8, 2005; final manuscript received September 15, 2006. Review conducted by Marc P. Mignolet.

the technology and apply the methodology to materials with more frequent defects such as powder nickel and materials with surface defects [6].

The zone-based methodology is clearly dependent on the discretization of the material into zones. The purpose of this research is to provide an extension to the probabilistic algorithm outlined in the AC to ensure that a converged risk solution is obtained regardless of the initial zone discretization. This is accomplished through the application of an iterative discretization refinement methodology akin to the "h" refinement method in finite element analysis, with critical components being: the criteria for determining regions of discretization, efficiently developing a refined model, and efficiently calculating an updated solution. The methodology is demonstrated through the analysis of an impeller disk.

Methodology Overview: A Zone-Based Approach

Probabilistic Methodology. Oftentimes, the critical locations of a structure are known a priori; typically, these locations are the highly stressed areas. However, since metallurgical defects are induced during the material melting process, defects may be randomly distributed within a disk. Thus, the location of a defect is a random variable. This fact considerably complicates a probabilistic analysis of a titanium structural component.

The random location of a defect in the cross section could be simulated using Monte Carlo sampling, that is, by randomly distributing the defect locations; however, an acceptance-rejection method would most likely be required since a disk cross section is irregular. One difficulty with this approach is that a large number of samples, e.g., several to many millions, would be required to ensure that all significant regions of the disk are sampled. For example, a highly stressed region near a notch would have a significant risk contribution but only a small volume and, therefore, may not be sampled sufficiently since the probability of locating samples within that region would be small. A second difficulty is that for each location the stress history must be obtained, typically from a finite element (FE) model, and processed to obtain the damaging fatigue cycles, i.e., rainflow cycle counting [7]. This operation would be very computationally demanding if carried out at every sample location.

An alternate approach is to adopt a "stratified sampling-like" method whereby the user divides the domain into regions (hereafter called "zones"), and sampling is performed in each region. This insures that every region defined by the user will be sampled. Thus, the user can tailor the sampling based on a priori knowledge about the likely high-risk regions of the disk. This leads to an efficient probabilistic approach; however, this leaves the accuracy of the result dependent on the user's definition of zones.

The second component of the alternate approach is to locate all defects placed within a zone at a single location in the zone and the cycles-to-failure and risk results for that location are assumed representative of a defect located anywhere in the zone. Thus, the stress information is retrieved and processed only once for all samples within a zone, resulting in a considerable savings in computer time. The location recommended in the AC is the life-limiting location of the zone, i.e., the location in the zone from which the cycles to failure will be the lowest. If the defect is located in the life-limiting location, the risk computed will be a conservative estimate of the risk for the zone.

The selection of the life-limiting location in a zone is not known a priori but can be determined by performing deterministic fatigue calculations for several candidate locations. In general, the stress and the proximity to a structural boundary are the two dominant considerations.

The methodology assumes a low frequency of anomalies; at most, only one significant anomaly exists in the disk; hence, the probability of two or more anomalies in a disk is negligible. This is consistent with the frequency of hard α defects in rotor-grade titanium [4,8]. For materials with more frequent anomalies (e.g., powder nickel), a Poisson distribution can be used to model the probability of occurrence with subsequent modifications in the probability equations [3,6,9].

The zone-based methodology accounts for:

- the probability of having an anomaly in the disk
- the probability that a hard α anomaly developed during the titanium melt process could be in any location of the disk
- the initial size distribution of the anomaly
- randomness in the time of inspection time, probability of detection, finite element stresses and fracture mechanics analysis
- the probability of fracture if an anomaly exists
- the probability of detecting an anomaly before the disk has fractured

The probability of fracture of the zones are modeled as independent events. The probability of fracture of the disk can be obtained from the system reliability equation

$$P[\text{disk}] = P[\text{fracture in any zone}] = P[F_1 \cup F_2 \cup \cdots \cup F_N] = 1 - \prod_{i=1}^{n} (1 - P[F_i]) \approx \sum_{i=1}^{n} P[F_i] \quad (1)$$

where F_i indicates fracture of a defect originating in zone i, $P[F_i]$ indicates the probability of fracture of an initial defect located in zone i, N is the total number of zones, $\cup$ is the union logical operator, and the approximate equality occurs because $P[F_i]$ is small.

The probability of fracture of a zone, i.e., $P[F_i]$ requires an anomaly be present and grow to failure. This can be represented as

$$P[F_i] = P[A_i]P[B_i|A_i] \quad (2)$$

where $P[A_i]$ is the probability of having an anomaly in zone i and $P[B_i|A_i]$ is the conditional probability of fracture given an anomaly in zone i.

The probability of a hard α defect being present per volume of material is dependent on the procedure used during the melting of the titanium and the inspection procedures used on the billet. Industry-developed exceedance curves define the frequency of occurrence and size of hard α defects [8]. The probability of having a defect in a zone, i.e., $P[A_i]$, is a function of the ratio of the zone volume to the volume of material used as a reference in developing the exceedance curve.

The conditional probability $P[B_i|A_i]$ is computed using probabilistic fracture-mechanics-based life assessment for low-cycle fatigue. The hard α phase anomalies found in titanium materials are typically brittle [4,8], and may form into growing cracks as early as the first cycle of applied load. For conservative fracture mechanics predictions, it is assumed that a crack forms at cycle zero, i.e., the crack initiation time is zero. Therefore, the cycles to failure is estimated exclusively using a crack propagation algorithm. The effects of inspections on the conditional POF can be determined by altering the conditional probability through the simulation of in-service inspections using anomaly detection and removal calculations based on crack size distributions and probability-of-detection curves [10]. The conditional probability calculations are computed using Monte Carlo sampling or importance sampling [11].

Fracture Mechanics Model. Evaluation of fatigue life requires calculation of stress intensity factors (SIFs) as a function of crack size. Calculation of SIFs for an arbitrary crack and structural geometry and stresses using finite or boundary element formulations are possible but not practical for probabilistic fatigue calculations. A practical yet reasonably accurate approach is to use weight functions [12,13]. Weight function methods for SIF determination require as input, the stress normal to the crack plane and the

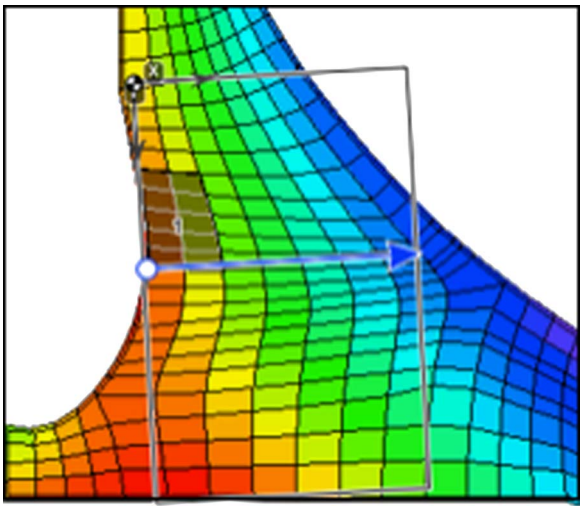


Fig. 1 Fracture mechanics model using a rectangular plate overlaid on an axisymmetric finite element model

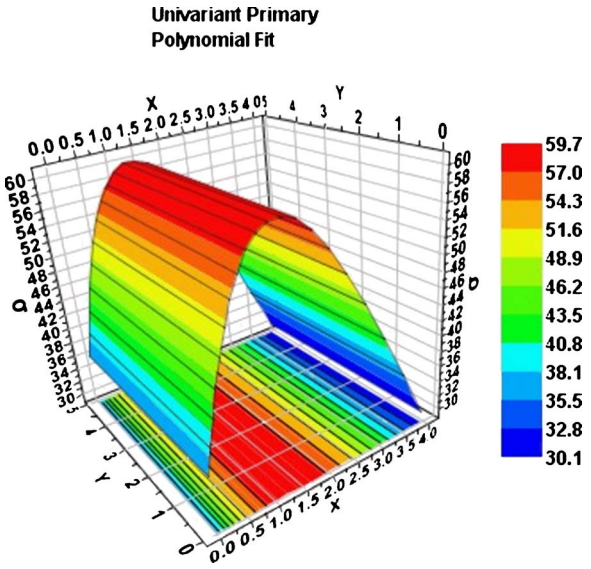


Fig. 2 Univariant stress gradient

geometric boundaries of the *uncracked* body. This allows one finite element solution to be used for all crack sizes. Univariant gradients and sometimes bivariant gradients in the stress field can be accommodated.

Weight functions have been developed for rectangular plate geometries and used as a surrogate for the more complex actual geometry [14]. This approximation is often sufficient because the SIF is only weakly affected by the boundary until the boundary is very close. During modeling, for each zone, the user defines a rectangular plate relative to the crack location and the boundaries of the actual part. Figure 1 shows the crack (white circle), the zone material (highlighted elements), the rectangular plate model (boundaries marked by black line), and the stress gradient direction (blue arrow), overlaid on a finite element model that is color coded based on the value of tangential stress. The crack is placed in the highest stressed location of the zone material, which is estimated to be the life-limiting location.

Selection of a rectangular plate that can adequately approximate the disk boundaries is subjective. A conservative approach can be chosen by making the plate boundaries smaller than the actual boundary and thus calculating a larger SIF and, hence, lower life estimate. Since the SIF is only weakly affected by the boundary proximity, approximating the boundary near the crack is most important.

The "gradient" vector shown in Fig. 1 indicates the direction along which the stress normal to the fracture plane is extracted from the FE results and used to construct a univariant gradient stress field normal to the fracture plane, as shown in Fig. 2. The vector does not restrict crack growth in any direction but only defines the direction in which a stress gradient will be considered when determining the stress intensity factor.

The gradient vector must not extend beyond the edges of the FE model in order to extract stresses along the vector. As such, a "clipping" algorithm is implemented within the graphical user interface to ensure this condition is satisfied; that is, the intersections of the vector and the disk boundaries are computed and, if necessary, the ends of the plate are automatically adjusted to ensure the vector lies within the disk geometry. The user has freedom to define the plate orientation and dimensions with the one restriction that the entire gradient vector must lie entirely within the FE model.

Random Variables. The random variables considered in the analysis are shown in Table 1. The anomaly location is addressed by the zone methodology discussed previously. The anomaly distribution defines the probability of having an anomaly and the size distribution of the anomalies. The FAA advisory circular provides industry-developed anomaly distributions based on melting practice and forging inspection processes [8]. The Aerospace Industries Association, Rotor Integrity Subcommittee, periodically updates the anomaly distributions. Variations in time of inspection are used to model real-world uncertainties in the inspection of a fleet of engines. The probability of detection defines the expected probability of detecting an anomaly as a function of anomaly size and inspection method.

Variations in finite element stress results are simulated using a multiplier that can be considered random, i.e.,

$$\sigma = \sigma(\mathrm{FE})S \tag{3}$$

where σ(FE) are the stresses obtained from the finite element analysis and S is a random variable modeled with a lognormal distribution.

A life scatter factor is implemented to consider variations in predicted cycles to failure, i.e.,

$$N = N(FM)B \tag{4}$$

where $N(FM)$ is the predicted cycles to failure from fracture mechanics analysis and B is a random variable modeled with a lognormal distribution. The life-scatter random variable encompasses all forms of variation in the computed cycles to failure with respect to the actual cycles to failure, such as material crack growth rate variation, stress intensity factor idealizations, etc.

Recent work has expanded the random variables that can be considered through application of a generalized conditional expectation method [15,16]. Thus, the zone-based methodology can be utilized with additional random variables such as geometric, loading, and material properties.

Convergent Zone Refinement Methodology

The risk solution obtained from the zone-based methodology is dependent on the discretization of the risk zones, i.e., the size and

Table 1 Random variables considered

Anomaly location
Anomaly distribution (size and frequency)
Time of inspection
Probability of detection
Finite element stress scatter
Life scatter (fracture mechanics multiplier)

Table 2 Zone refinement methodology

No.	Step details	Operator[a]
1	User defines initial zones and sets parameters—convergence criteria, maximum limits	U
2	Risk assessment evaluated	C
3	Examine stop criteria a. Risk $\leq$ risk limit, stop analysis b. Relative change in disk risk $<$ change limit, stop analysis c. All risk contribution factor (RCF) values $<$ RCF limit d. Maximum cpu time reached	U
4	Select zones to be refined based upon the zone risk contribution factor—with or without inspection a. Zone RCF $<$ RCF factor limit, no refinement b. Zone RCF $\geq$ RCF limit, refine into subzones	U
5	Generate potential subzones a. Determine material in each subzone i. Use stress centroid equation ii. Embedded—create four or three subzones iii. Surface—create two subzones b. Place flaw in each subzone geometrically closest to flaw in parent zone c. Define plate for each subzone Parent plate orientation, size and location used as default d. Inherit properties from parent Gradient direction, volume multiplier, inspection schedules, material no., crack type, crack plane, anomaly distribution, # samples	C
6	Read risk results from unchanged zones from database	C
7	Compute risk for new zones, combine with unaltered zones, that is, return to step 2	C

[a]U and C denote User and Computer, respectively.

number of zones. The methodology does, however, require that the flaw be placed in the life-limiting location of the zone; the calculated risk is, therefore, greater for coarse discretizations compared to fine zone discretizations. In other words, the risk should be reduced as the number of zones is increased.

A methodology is presented that provides a converged risk solution independent of the initial discretization of zones. The features of the algorithm are

- Robust—will work for any well-posed problem
- Convergent—will converge to correct solution
- Simple—easy to understand, not hidden nor confusing
- Stable—risk solution is not dependent on the initial zone breakup

The detailed zone refinement methodology is shown in Table 2. Steps 3–7 are iterated as necessary.

The iterative algorithm is implemented through a combination of user actions and computational algorithms as denoted in column 3 of Table 2 where "U" denotes a user action and "C" denotes a computational algorithm.

Selecting Zones to Refine. Identification of the zones requiring refinement (Step 4 in Table 2) is a key ingredient to the zone refinement algorithm. The computational time associated with prediction of disk risk is directly proportional to the number of zones. For optimum computational efficiency, zone refinement should be applied only to the zones that significantly influence the overall disk risk.

The risk contribution factor (RCF) for a zone indicates the risk contribution of the zone relative to the risk of the disk, given mathematically as

$$\mathrm{RCF}_i = \frac{P[F_i]}{\sum_{j=1}^{n} P[F_j]} = \frac{P[A_i]^* P[B_i|A_i]}{\sum_{j=1}^{n} P[A_j]^* P[B_j|A_j]} \tag{5}$$

where N is the total number of zones.

The risk contribution factors are defined such that

$$\sum_{i=1}^{n} \mathrm{RCF}_i = 1 \tag{6}$$

Therefore, a percent limit, (RCF limit, Table 2, Step 4), can be used to determine if a zone is significant and should be refined. Through this approach, critical zones are refined until all zones have an RCF less than an RCF limit or some other exit criterion is reached (see Step 3 in Table 2). Note, each zone has a separate RCF for analyses with or without inspections and these RCF values may not be the same. As a result, the largest RCF value from the with and without inspections results is typically used.

In Fig. 3, zones with an RCF $>5\%$ are highlighted in red (gray). These zones are candidates for zone refinement as described in Steps 4 and 5 of Table 2 and discussed below. The user can change the RCF threshold value as desired.

Zone Refinement

Partitioning Parent Zone into Subzones. The procedure to refine a zone is to first subdivide the zone area into subzones. This is shown schematically in Fig. 4. The parent zone, highlighted material on the left, is subdivided into fourths (sometimes thirds depending on the geometry) about the stress centroid of the zone

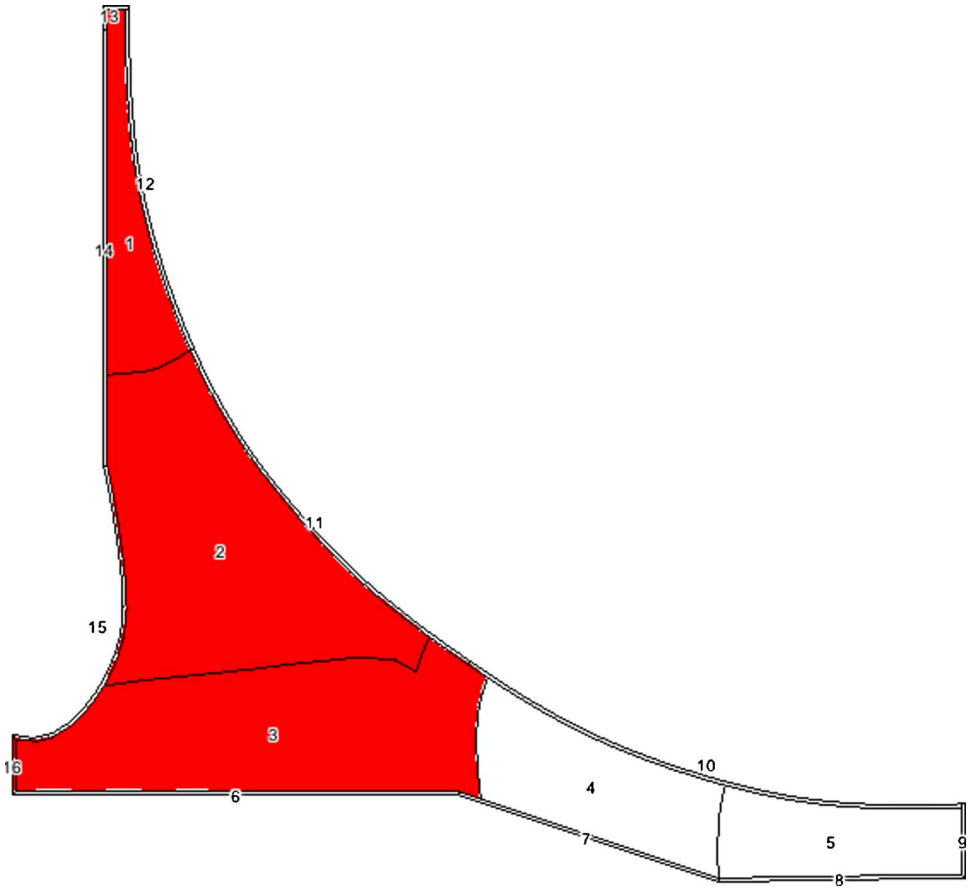


Fig. 3 Selection of zones for refinement using risk contribution factors; red (gray) zones indicate RCF greater than 5%

using the local coordinates of the plate. The stress centroid is defined by coordinates R_{cg} (radial), Z_{cg} (axial) as

$$R_{cg} = \frac{\sum_{k=1}^{K}\sum_{l=1}^{L} R_{lk}\sigma_{lk}}{\sum_{k=1}^{K}\sum_{l=1}^{L} \sigma_{lk}} \qquad Z_{cg} = \frac{\sum_{k=1}^{K}\sum_{l=1}^{L} Z_{lk}\sigma_{lk}}{\sum_{k=1}^{K}\sum_{l=1}^{L} \sigma_{lk}} \tag{7}$$

where K denotes the number of finite elements in the zone to be divided, L denotes the number of nodes within element k, R_{lk} and Z_{lk} are the radial and axial locations of node l within element k, and σ_{lk} is the stress at node l within element k. Typically, the user selects the most damaging stress component and load case on which to base refinement. The advantage of using the stress centroid equation instead of the geometric equation is that the centroid will be placed closer to the highly stressed region, ensuring smaller subsequent zones within that region.

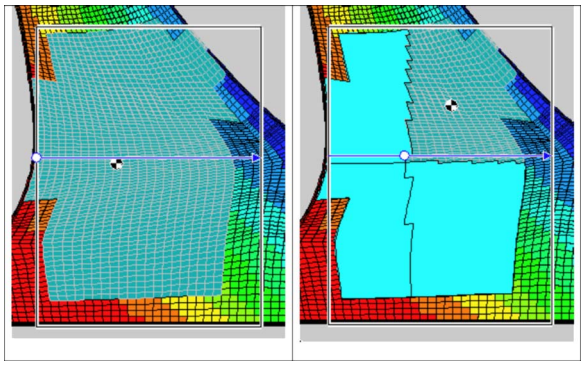


Fig. 4 Schematic of zone subdivision: parent zone (left) divided into four subzones (right) about stress centroid of parent

Once the centroid is located, the zone is bisected into two, three, or four subzones. For defects initiating in the interior of the part, i.e., embedded cracks, lines emanating from the zone's stress centroid parallel to the sides of the plate are used to determine the subzone location of each finite element based on the element's geometric centroid location relative to the lines. Generally four subzones are created, see Fig. 4; however, an oddly shaped parent zone, e.g., "L" shaped, may result in only three subzones. There is no restriction on the shape of the zones.

For defects initiating on the surface of the part, i.e., surface cracks, only a single row of elements along the surface comprises the zone. A line parallel to the side of the plate (usually normal to the surface) passing through the stress centroid is used to partition the parent zone into two subzones.

Defining Crack Location and Plate Geometry. Next, the flaw location must be specified for each subzone. For the parent zone, the flaw is by definition located in the life-limiting location. For the new subzone, it is theoretically possible to compute the life at multiple locations, say each node point within the zone, and then select the location with the lowest life. However, this solution is not executed by default due to excessive computation time. Another option is to locate the flaw at the highest stress in the zone based on the stress state from the most damaging load case. However, this option ignores the amplification effect near surfaces have on the stress intensity factor and, therefore, may not be conservative. The approach implemented is to locate the subzone flaw

within the subzone material at the geometrically closest point to the flaw location in the parent zone. This approach is predicated on the premise that the parent flaw is in the life-limiting location, which by definition, considers all factors affecting life, primarily stress level and location. The location of the flaw for the new zone is shown in the right image in Fig. 4 for the upper rightmost zone.

It is possible that the flaw location in the subzone is not the life-limiting location, particularly if a large zone is subdivided and other boundaries become important for the subzone or significant stress gradients occur within a parent zone. However, the zone-refinement scheme reduces the probability of not locating a defect in the life-limiting location because subdividing the zone reduces, typically, by a factor of four, the cross-sectional area associated with that zone, and the cycles-to-failure of a defect located anywhere in the zone becomes more homogeneous. Second, the zone-refinement scheme reduces the importance of determining the life-limiting location for any particular zone because any zone of significance is, by definition, subdivided until the RCF is below a limiting value. As such, any individual zone cannot have a dominant effect on the overall disk risk, and thus, not placing the defect in the life-limiting location in any particular zone will not grossly affect the overall disk risk calculation.

Next, a rectangular plate must be defined for the subzone for the fracture mechanics solutions. The procedure is to:

1. Use the same plate as the parent zone (assuming the new crack is inside the existing plate), and keep the same gradient direction.
2. Clip front and back along gradient line if necessary.
3. If the new flaw location is outside parent plate, then move the plate if possible. If not possible, then warn the user so that the late may be repositioning manually.

Other properties necessary for the risk analysis are inherited from the parent zone, such as material properties, inspection schedules, and anomaly distribution.

Efficient Reanalysis—Computational Issues. After subdivision of selected zones, the model is reanalyzed to determine the updated risk and RCFs. Several, if not many, of the zones will not have been altered and, therefore, have the same risk value as before. In support of reanalysis, an XML results database is used to store results by zone from a previous analysis [17]. During re-analysis, for each zone, the database is scanned in search of a matching zone. If a match is found, then the results are retrieved and analysis proceeds to the next zone. If a match is not found, then the POF for the zone is computed and the results are stored in a new database. This capability provides a large reduction in computer time versus rerunning all zones.

The zone-based methodology is defined such that the determination of the probability of fracture for the zones can be computed independently. A parallel processing spatial decomposition methodology using a network of workstations and personal computers is implemented whereby each computer analyzes a region of the disk, i.e., a single zone or a collection of zones [18]. Each analysis creates an XML database. The individual analyses are then combined by searching across all databases to obtain the probability of fracture of the disk. This methodology produces an order of magnitude or more improvement in computation time.

Supporting Technology: Element Refinement and Onion Skinning. The finite element mesh is used as a framework for determining zone boundaries. However, although atypical, situations may arise where the FE mesh is too coarse for the risk assessment. This may occur, for example, near surfaces where there is little or no stress gradient but a significant risk gradient due to the amplification effect of the near surface on the SIF. For these situations, an "element refinement" option can be used to repeatedly subdivide the FE mesh. Any nodes created are interior to existing elements; therefore, stress components for any new nodes are obtained by interpolating the stress components from the original mesh using the element shape functions.

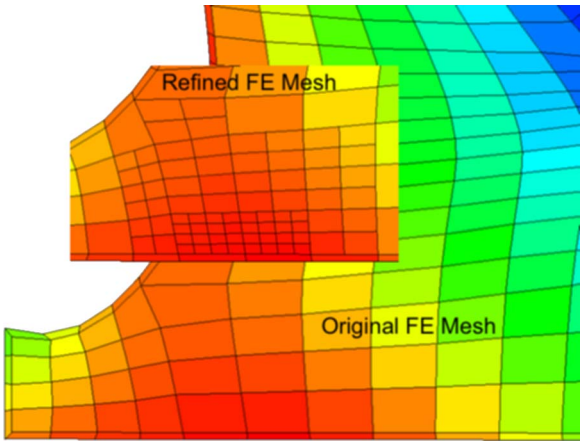


Fig. 5 Example of finite element subdivision in support of zone refinement

Figure 5 shows an example of element subdivision of a region near a surface. The FE mesh is only used to define zone boundaries; it is not used for FE analysis. Thus, there are no continuity requirements along element edges and any localized subdivision is allowed. It is assumed that the finite element model is sufficiently refined with respect to stress gradients; therefore, subdividing the elements and interpolating the stress results using element shape functions does not introduce error.

The advisory circular states that a 20 mil (0.020 in.) layer, e.g., onion skin, of material be associated with surface flaws. However, the FE mesh was most likely developed without this requirement in mind. Therefore, an algorithm has been developed and implemented to automatically create a layer of elements of the specified thickness.

The onion skinning algorithm computes a vector of the required length, e.g., δ, normal to the element edge. If the projected distance to the normal of the nearest node to the element edge is $<\delta$, no subdivision is required. If the projected distance is $>\delta$, additional nodes are located along the current element edge such that the perpendicular distance to the surface is δ. The original element is then split into two elements. Figure 6 shows an example of onion skinning. The "red" (dark gray) nodes indicate the nodes of the original element. The "green" (gray) nodes and green (gray) element side denote components generated by the onion-skinning algorithm. Elements "a" and "b" are created and the original element is deleted. Stress components and temperatures are interpolated for the new nodal points. The onion-skinning algorithm is summarized in Table 3.

Exit Criteria. The exit criteria used to terminate the iterative algorithm vary. The first criterion to examine is whether the risk

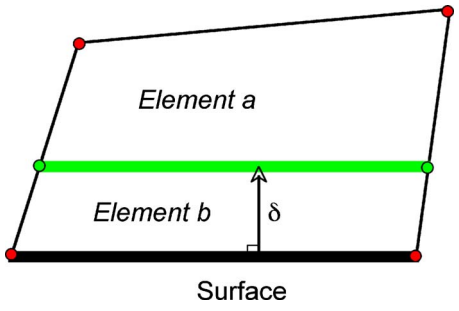


Fig. 6 Example of onion-skinning algorithm applied to a single finite element

Table 3 Onion skinning algorithm

1	User graphically selects set of elements to consider for onion skinning
2	Filter out nonsurface elements: a. Compute geometry border as set of edges b. Iterate through elements checking for an element edge in border set
3	For each selected border element: a. Create a vector normal to the surface edge with length equal to the onion skin distance δ b. Using the dot product, determine the distance along element edge that gives a normal distance to the surface equal to δ
4	Using linear interpolation between adjacent edge endpoints and the normal distance δ, compute location, temperature and stress components for new nodes
5	Delete old element and replace with two new elements using new nodes

limit, i.e., the DTR, has been satisfied. If the life-limiting location has been chosen for each zone, then the POF estimate is conservative and further zone discretization will reduce the POF. Therefore, if the POF is below the risk limit, then the design is sufficient and the algorithm can be terminated.

A second exit criterion is to examine the RCF values for all zones. If the RCF values are all below a reasonable RCF threshold, then further zone discretizations will not likely produce a significant change in the risk. An RCF threshold of 1-5% is typical. For example, if the RCF for a zone is 2%, then subdividing that zone will at most affect only 2% of the total disk risk. Summation of the risk in all zones above the RCF limit gives an indication of the total amount of disk risk that might be affected by zone refinement.

A related criterion is to examine the relative change in the risk between iterations. If the change in risk is small, then further zone discretizations will not likely produce a significant change in the risk.

A final criterion is to examine the maximum limits for the analysis, e.g., maximum number of iterations or the maximum cpu time is reached. These are practical criteria based mainly on experience.

If the risk is greater than the DTR, further refinement can be attempted to meet the target. If the target cannot be obtained, then disk redesign is required.

Numerical Example

An impeller model is used to demonstrate the algorithm. The three-dimensional model is shown in Fig. 7. Because of axisymmetry, only the cross-sectional area is discretized for finite element and risk analysis, as shown subsequently. The procedure outlined in Table 2 is performed using an RCF threshold of 5%. Any zone with an RCF of $\geq$5% for either with or without inspection is subdivided. The iterative procedure is carried out until the RCF values for all zones are $<$5%.

A refinement sequence is shown in Figs. 8(*a*)–8(*f*), starting from a very coarse 16 zone model (iteration 0) and concluding, after iteration 6, with a refined 75 zone model. (Note, there are several very thin surface zones around the boundary of the model that cannot be seen at this scale.) The zones with an RCF greater or equal to 5% are colored red (gray) and others white. The red (gray) zones are refined in the succeeding iteration.

In each iteration, the current number of total zones in the model is indicated along with the number of zones that are unchanged and, hence, retrieved from the results database. The percent time for solving that particular model relative to the total computational time for all iterations is given. Also given is the cumulative percentage of time from iteration zero. For example, the results after iteration 1 show a total of 25 zones with 13 unchanged from

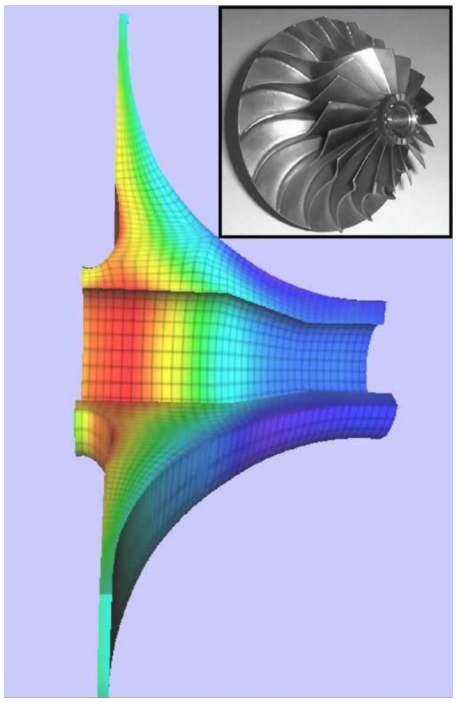


Fig. 7 Three-dimensional idealized impeller finite element model

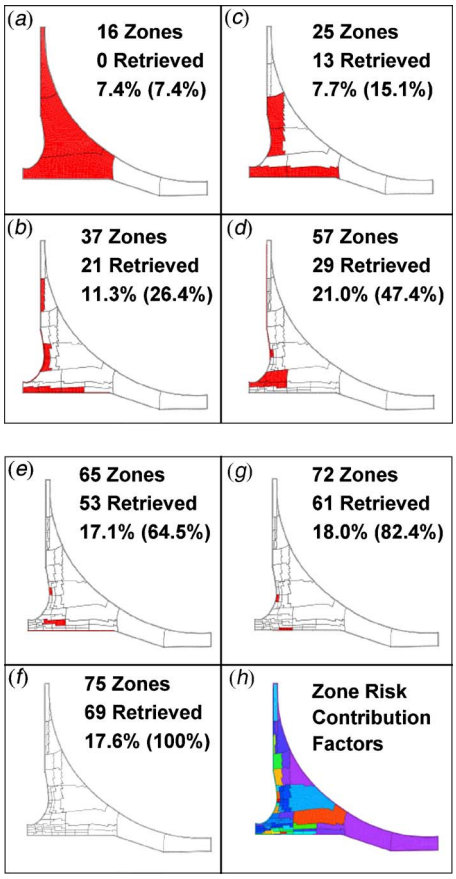


Fig. 8 (a)–(d) Sequence of zone refinement iterations for impeller rotor disk model and (e)–(h) sequence of zone refinement iterations for impeller rotor disk model

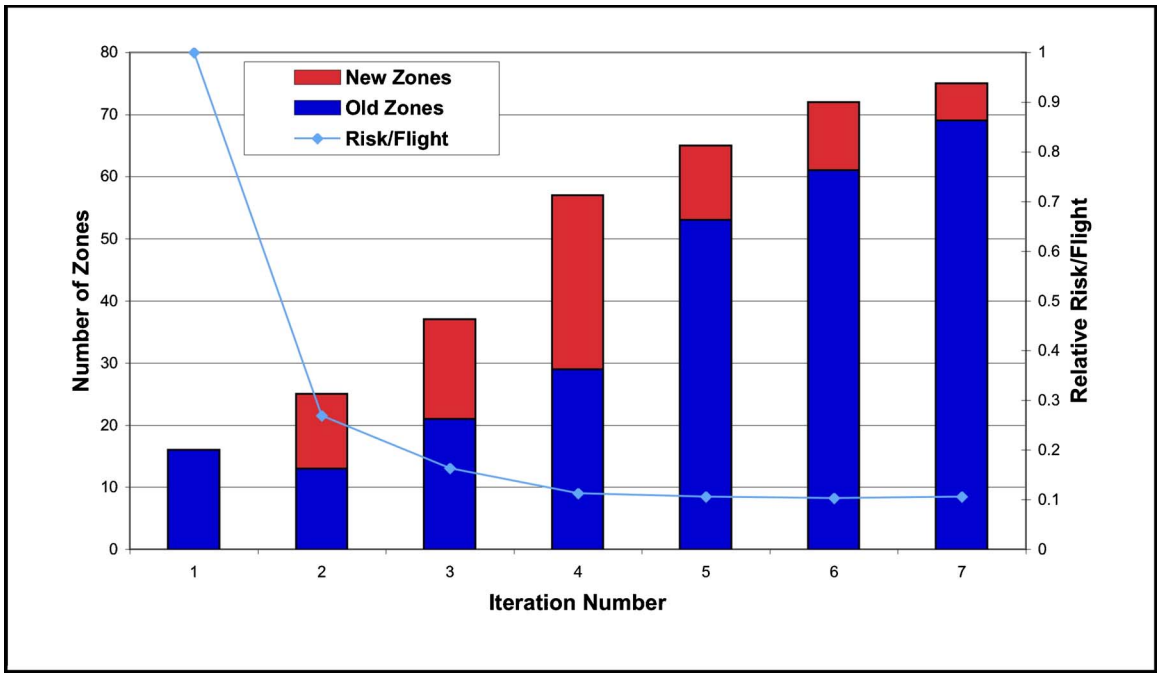


Fig. 9 Iteration results showing the convergence in risk versus the number of zones (old and new)

iteration 0. The solution time for iteration 1 is 7.7% of the total computational time. The cumulative solution time for iterations 0 and 1 is 15.1% of the total computational time.

Figure 8(*g*) shows the converged solution with the zones colored using a color map from high RCF (red) to low RCF (purple); however, all zones have an RCF of $<5\%$.

Figure 9 shows a chart of the normalized POF (normalized to the initial computed risk) versus the iteration number. Since the life-limiting location is chosen for the defect in each zone, the calculated POF is conservative and subsequent refinement reduces the POF. The results are typical; a large reduction in risk occurs over the first few iterations with smaller changes in subsequent iterations. The percent change in POF between iterations zero and one, one and two, etc. are 73, 39, 31, 6, 2, 2, respectively. Thus, after iteration four, the percent change in overall POF between iterations is reduced to 2%. The final POF is approximately 11% of the POF from iteration 0.

The red (gray) portion of the column shows the number of new zones created during that iteration and the "blue" portion shows the number of zones retrieved from the previous iteration. The number of zones retrieved is always less than the total number of zones from the previous iteration since parent zones are deleted after being subdivided.

Conclusions

A probabilistically based damage tolerance methodology has been developed in concordance with the Federal Aviation Administration Advisory Circular 33.14-1 to compute the probability of fracture of gas turbine titanium disks subject to low-frequency metallurgical (hard α) defects. This methodology is used during design to augment the traditional safe-life approach and ensure that the risk of fracture due to hard α defects is below a designated design target risk.

This research extends the probabilistically based damage tolerance methodology to ensure that a converged probability of fracture result is obtained that is independent of discretization issues and the experience of the analyst.

The methodology is robust, simple to understand, will converge to the correct solution, and is only weakly dependent on the initial zone discretization. The technique used is akin to "h" refinement in finite element analysis in that a local metric, the risk contribution factor, is used to locate the zones requiring further discretization. These zones are then subdivided about the stress centroid of the zone. Risk results for the new zones are then obtained and combined with the risk from the existing zones.

Efficiency issues are addressed through the use of an automated zone refinement algorithm, an XML-based results database and the parallel processing of zone calculations.

A numerical example of the risk assessment of an impeller is presented. The results are representative of a general application of the algorithm to titanium disks. Four to six iterations are typically sufficient to obtain a reduction in risk of approximately an order of magnitude.

Acknowledgment

This work was performed as part of the Turbine Rotor Materials Design (TRMD) project supported by the Federal Aviation Administration under Cooperative Agreement No. 95-G-041 and Grant No. 99-G-016. The authors would like to acknowledge the collective contributions of the TRMD program team: Southwest Research Institute, General Electric, Honeywell, Pratt & Whitney, Rolls-Royce, the University of Texas at San Antonio, and Mustard Seed Software.

Nomenclature

AC = Federal Aviation Administration Advisory Circular
B = lognormal distribution random variable applied to computed cycles-to-failure
DTR = design target risk
F_i = fracture event of zone i
FAA = Federal Aviation Administration
FE = finite element method
FM = fracture mechanics
hard α = brittle inclusion possibly developed during titanium material melting
life-limiting location = location in the zone with the lowest cycles to failure

n = number of zones
N = cycles to failure
$N(FM)$ = computed cycles to failure from fracture mechanics model
onion skinning = algorithm used to develop a thin layer of elements surrounding the surface
$P[A_i]$ = probability of an anomaly being present in zone i
$P[B_i|A_i]$ = probability of fracture of zone i given an anomaly being present
$P[F_i]$ = probability of fracture event occurring in zone i
POF = Probability of Fracture
R_{cg} = radial coordinate of zone stress centroid
R_{lk} = radial coordinate of node l within element k
RCF = risk contribution factor
S = lognormal distribution random variable applied to driving stress
SIF = stress intensity factor
Z_{cg} = axial coordinate of zone stress centroid
Z_{lk} = axial coordinate of node l within element k
zones = regions of similar cycles to failure
σ = driving stress applied to crack
σ(FE) = driving stress obtained from finite element model
σ_{lk} = hoop stress of node l within element k
δ = distance normal to surface used for onion skinning algorithm

References

[1] NTSB, 1990, "Aircraft Accident Report—United Airlines Flight 232 McDonnell Douglas DC-10-10 Sioux Gateway Airport, Sioux City, Iowa, July 19, 1989," NTSB/AAR-90/06, National Transportation Safety Board, Washington, DC.

[2] Federal Aviation Administration, 2001, Advisory Circular 33.14-1: "Damage Tolerance for High Energy Turbine Engine Rotors," U.S. Department of Transportation, AC, Washington, DC.

[3] Roth, P. G., 1999, "Probabilistic Rotor Design System (PRDS), Gas Turbine Engine Design," Propulsion Directorate, Air Force Research Laboratory, Wright-Patterson Air Force Base, ARFL-PR-WP-TR-1999-2122.

[4] Southwest Research Institute, Allied Signal, General Electric, Pratt and Whitney, Rolls-Royce Allison, Scientific Forming Technologies, 2000, "Turbine Rotor Material Design—Final Report," Federal Aviation Administration, Washington, DC, DOT/FAA/AR-00/64.

[5] Leverant, G. R., Millwater, H. R., McClung, R. C., and Enright, M. P., 2004, "A New Tool for Design and Certification of Aircraft Turbine Rotors," ASME J. Eng. Gas Turbines Power, **126**(1), pp. 155–159.

[6] Enright, M. P., and Huyse, L. J., 2005, "Methodology for Probabilistic Life Prediction of Multiple Anomaly Materials," AIAA J., **44**(4), pp. 787–793.

[7] Downing, S. D., and Socie, D. F., 1982, "Simplified Rainflow Counting Algorithm," Int. J. Fatigue, **4**(1), pp. 31–40.

[8] Subteam to the Aerospace Industries Association Rotor Integrity Subcommittee, 1997, "The Development of Anomaly Distributions for Aircraft Engine Titanium Disk Alloys," *38th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Kissimmee, FL, pp. 2543–2553.

[9] Ang, A. H.-S., and Tang, W. H., 1975, *Probability Concepts in Engineering Planning and Design*, Vol. 1, Wiley, New York.

[10] Wu, Y.-T., Enright, M. P., and Millwater, H. R., 2002, "Probabilistic Methods for Design Assessment of Reliability With INspection," AIAA J., **40**(5), pp. 937–946.

[11] Huyse, L. J., and Enright, M. P., 2003, "Efficient Statistical Analysis of Failure Risk in Engine Rotor Disks Using Importance Sampling Techniques," *44th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Norfolk, VA.

[12] Buecker, H. F., 1970, "A Novel Principle for the Computation of Stress Intensity Factors," Z. Angew. Math. Mech., **50**, pp. 529–546.

[13] Wu, X.-R., and Carlsson, A. J., 1991, *Weight Functions and Stress Intensity Factor Solutions*, Pergamon Press, London.

[14] McClung, R. C., Enright, M. P., Lee, Y.-D., Huyse, L. J., and Fitch, S. H. K., 2004, "Efficient Fracture Design for Complex Turbine Engine Components," ASME Paper No. GT2004-53323.

[15] Momin, F. N., Millwater, H. R., Osborn, R. W., and Enright, M. P., 2004, "Application of a Conditional Expectation Response Surface Approach to Probabilistic Fatigue," *9th ASCE Specialty Conference on Probabilistic Mechanics and Structural Reliability*, Albuquerque.

[16] Momin, F. N., 2004, MS thesis "Application of the Generalized Conditional Expectation Method to Enhance a Probabilistic Design Code," The University of Texas at San Antonio.

[17] Millwater, H. R., Fitch, S. H. K., Enright, M. P., and Huyse, L., 2003, "Application of an XML-Based Database for Probabilistic Analysis," AIAA Paper No. 2003-1837.

[18] Millwater, H. R., Shook, B. D., Guduru, S., and Constantinides, G., 2004, "Application of Parallel Processing to Probabilistic Fracture Mechanics Analysis of Gas Turbine Disks," AIAA Paper No. 2004-1745.

Confidence Interval Simulation for Systems of Random Variables

Thomas A. Cruse[1]
Chief Technologist (IPA)
Air Force Research Laboratory,
Wright-Patterson AFB, OH 45463
e-mail: Thomas.Cruse@wpafb.af.mil

Jeffrey M. Brown
Structural Modeling Engineer
Air Force Research Laboratory,
Propulsion Directorate,
Wright-Patterson AFB, OH 45463
e-mail: Jeffrey.Brown@wpafb.af.mil

Bayesian network models are seen as important tools in probabilistic design assessment for complex systems. Such network models for system reliability analysis provide a single probability of failure value whether the experimental data used to model the random variables in the problem are perfectly known or derive from limited experimental data. The values of the probability of failure for each of those two cases are not the same, of course, but the point is that there is no way to derive a Bayesian type of confidence interval from such reliability network models. Bayesian confidence (or belief) intervals for a probability of failure are needed for complex system problems in order to extract information on which random variables are dominant, not just for the expected probability of failure but also for some upper bound, such as for a 95% confidence upper bound. We believe that such confidence bounds on the probability of failure will be needed for certifying turbine engine components and systems based on probabilistic design methods. This paper reports on a proposed use of a two-step Bayesian network modeling strategy that provides a full cumulative distribution function for the probability of failure, conditioned by the experimental evidence for the selected random variables. The example is based on a hypothetical high-cycle fatigue design problem for a transport aircraft engine application. [DOI: 10.1115/1.2718217]

Motivation

Probabilistic design is becoming recognized within the turbine engine community as a major requirement for advanced Air Force propulsion systems where weight and cost are critical component drivers. Probabilistic design provides a physics-based assessment of what the expected component and system performance will be, including overspeed capability of turbine disks, low-cycle fatigue, and high-cycle fatigue (HCF). Furthermore, probabilistic design provides the analyst with sensitivity factors for performance and cost that provide direct insight into the key drivers for unreliability or reliability-driven cost factors.

There are two critical issues that we believe have to be addressed before probabilistic design methods can be applied to component designs for advanced propulsion systems for military or commercial applications—validation of the design methods and a formal assessment of confidence in the design predictions. The issue of validating probabilistic design is the subject of another study. This paper reports on a proposed prediction methodology for computing confidence intervals on the results of Bayesian network models used for probabilistic system performance simulation.

Confidence interval (CIs) are either frequentist or Bayesian. In statistical analysis, the CI is established by determining the value of a probability distribution parameter that is the best estimate of the value that matches a finite set of available data. The method of maximum likelihood parameter estimation typifies the statistical approach. The number of data points strictly drives the width of the CI in the experimental data set.

The Bayesian approach to defining a belief or confidence interval is not driven by a single, optimal value of a parameter, but is rather derived from conditional probability distribution that is derived from a belief that the parameter of interest is described by a probability distribution of its own. Of course, we may have data to use in estimating the distribution of the parameter, resulting again in a CI driven by the number of data points.

Bayes' theorem may be used to combine prior knowledge, such as expert opinion estimation of the parameter uncertainties, with limited evidence to obtain a CI for the posterior distribution that combines both expertise (prior experience) and data. This combination of information is unique to the Bayesian CI and best represents what we see as the likely design environment for probabilistic design of turbine engine components, particularly in the early stages of design where data are more limited.

The value of properly computed Bayesian CIs for design lies in their ability to tie the estimated CI to specific physical probabilistic sensitivity factors. Such probabilistic sensitivity factors combine the physical sensitivity of a given variable with the uncertainty of the variable. When done in this way, the designer has the ability to determine the most cost-effective experimental strategy for improving the CI (reducing the range of the interval).

Many see Bayesian network models[2] for complex systems to be a very useful modeling environment for probabilistic turbine engine design problems. The features of Bayesian networks that are very attractive for design include the following items:

- The graphical network (based on directed acyclic graphs (DAGs) provides an unambiguous causality model for the physics-based design models.
- The network nodes allow for a wide variety of response surface or reduced-order model approximations to be combined to achieve multiphysics probabilistic models governing component behavior.
- The efficiency for large system probabilistic simulation that is provided by the Markov chain Monte Carlo (MCMC) algorithms used for the solution of Bayesian networks.
- The ability to blend evidence and expertise through Bayesian updates of the top-level, data-driven nodes in the network.
- The ability to meld evidence for any number of the top-level, data-driven nodes in the probabilistic simulation.

[1]Corresponding author. Professor of Mechanical Engineering, Vanderbilt University.

Submitted to the ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received September 16, 2005; final manuscript received October 11, 2005. Review conducted by Lee Langston.

[2]Note that we are using Bayesian terms in two different modes in this study: one is the notion of a Bayesian network where Markov chain properties derive from the direct use of Bayes' theorem; second is the use of prior estimates for random variable distribution parameters that are subsequently updated to posterior distributions through another application of Bayes' theorem.

However, we have determined that Bayesian networks do not support the determination of Bayesian CIs needed for the design-assessment requirement we believe is critical for both military and commercial certification. The purpose of this paper is to demonstrate this deficiency in Bayesian network modeling and to provide an effective work-around to the limitation that maintains all of the above benefits of Bayesian network modeling.

Conditional Probability Integral and Confidence Intervals

We define two independent random variables S and R for our particular application. The first of these RVs is defined as a function of independent random design variables X_i conditioned on a vector of parameters $\vec{\vartheta}$, which are associated with the probability distribution for S. It must be noted that each of the random design variables X_i has its own individual probability distribution that generates the joint probability distribution for S based on the problem physics. The vector $\vec{\vartheta}$ defines the conditioning parameters, and each has its own independent probability distributions.

The RV $Z=S-R=0$ defines a conditional failure limit state for this problem dependent on the values of the probability distribution parameters. This limit state defines the failure region for all realizations of the RVs where $S>R$. Our purpose is to estimate the $\Pr(Z>0)$ given as the following conditional probability of failure.

Before giving the integral form for the estimated probability of failure, a few words of clarification are needed. In theory, one can consider that there exists for any system a true probability of failure (a number) that is not dependent on (conditional) uncertain knowledge. The purpose of the current work is to estimate this true value by a distribution of the probability of failure, given that we have imperfect or incomplete knowledge of the conditional dependencies of the probability distributions. Thus, for our purposes herein, the probability of failure is a random variable conditioned by hypothetical or experimentally based conditioning parameters $\vec{\vartheta}$ and $\vec{\varphi}$, as follows:

$$\Pr_F(\vec{\vartheta},\vec{\varphi}) = \Pr(Z>0|\vec{\vartheta},\vec{\varphi}) = \int_0^\infty f_z(z;\vec{\vartheta},\vec{\varphi})dz \tag{1}$$

We will refer to Eq. (1) as the "inner loop" for a typical Monte Carlo simulation associated with the probability of failure, conditionally based on the set of parameter vectors $\vec{\vartheta}$ and $\vec{\varphi}$ for the two RVs. In general, the conditioning parameters are themselves random variables (uncertain parameters). Their influence on the probability of failure will be captured in an "outer-loop" simulation. In general, the probability of failure probability distribution cannot be computed in an analytical form and is estimated using a two-step Monte Carlo simulation process.

The probability of failure written for the α^* level of confidence (in percent) is conceptually obtained by computing Eq. (1) for all values of the conditioning parameters subject to the constraint defined by the selected value of the parameter α^*. The value of α^* is a percentage that corresponds to a selected confidence interval level (CI) as notionally given below, for all values of the parameters $\vec{\vartheta},\vec{\varphi}$

$$\Pr_F(\alpha^*) = \Pr[\Pr_F(\vec{\vartheta},\vec{\varphi}) \le \alpha^*] \tag{2}$$

We adopt the notation that the implied probability integration in Eq. (2) over the range of conditional parameter probability density functions (PDFs) is the "outer-loop" simulation for the probability of failure. This is called an outer-loop simulation as the value of the $\Pr_F(\vec{\vartheta},\vec{\varphi})$ must be computed first for each parameter value (the inner loop). The purpose of the current report is to present an approximate way to compute the outer-loop integration.

To perform the outer-loop probability integration, we use a response surface (RS) for the conditional probability given in Eq. (1). Given the product nature assumed by parameter distribution independence (one will have to revert to the physics of the problem to validate this assumption), we take a relatively small design of experiments (DOE) definition to get the needed RS. Such an inner-loop RS model is then used in an efficient algorithm for large problems, such as the Markov chain Monte Carlo (MCMC) algorithm. This approach appears to be easily extended to include multiple random distributions when those are also taken to be uncertain. An example of the RS approach is reported herein.

Markov Chain Monte Carlo System Reliability Model for High-Cycle Fatigue (HCF)

It is neither the intent nor the scope of this paper to derive the details of the MCMC algorithm. We will use the Markov Chain Monte Carlo (MCMC) algorithm implementation available as a free MS/WINDOWS code known as WINBUGS [1]. The reasons for using the Bayesian network have already been elucidated. The application in this section will be seen to be a forward simulation only and not a full Bayesian updating analysis, using the MCMC algorithm. A full Bayesian treatment of the problem will be given in the next section.

The Bayesian network simulation algorithm is based on an iterative sampling strategy within a causality network of conditional probabilities that produces a stable Markov chain [2]. The network conditional PDFs are iteratively updated using Bayes' theorem for the linked network such that each iteration produces a posterior simulation value for each node based on the prior iteration. The algorithm results in a significant computational effort reduction over using non-network-based Monte Carlo algorithms.

Unfortunately, the Bayesian network algorithm does not directly support the simulation of the CDF of the probability of failure to determine the CI result that we seek. The reason is simply based on the fact that the Markov chain resulting from the network model for Eq. (2) samples combining both the inner- and outer-loop contributions of the nodal probability distributions with each iterative sweep of the network. As a result, we do not recover the independent effect of the uncertainties of the conditioning parameters on each, conditional probability of failure.

We propose to circumvent this limitation in Bayesian network modeling by approximating the inner-loop probability integral in Eq. (1) and imbedding this approximation in the Bayesian network for the outer-loop simulation. However, before we do that, we begin by setting up a representative design problem for a hypothetical high-cycle fatigue (HCF) limited design of a rotor stage. For this first example, the conditioning parameters will be taken as known rather than random variables. For this case, there is a single theoretical value of the probability of failure and not a probability distribution. The following section addresses uncertainties in the probability distribution parameters for the mechanical variables.

The hypothetical HCF problem considers a low-order blade resonance to be equivalent to blade failure, thereby ignoring material-based HCF damage models for this illustration. The simulation is intended to cover a total fleet of engines for which this failure can occur, albeit in the context of a single blade design for the example. The inner-loop problem is the portion comparing the stress and strength values for a given simulation, which values are simulated based on the conditioning parameters for the various mechanical variables.

The design problem considers the following highly simplified scenario. The engine/aircraft are taken to be in the transport category. The critical resonance design condition is a blade fundamental resonance mode taken to be above the operating speed range for the engine high-pressure (HP) rotor. During takeoff, the HP rotor speed increases and achieves its maximum speed for the mission. The takeoff is limited to 90 s maximum prior to throttle reduction for aircraft climb. The reliability of the system must consider the following random variables or design factors:

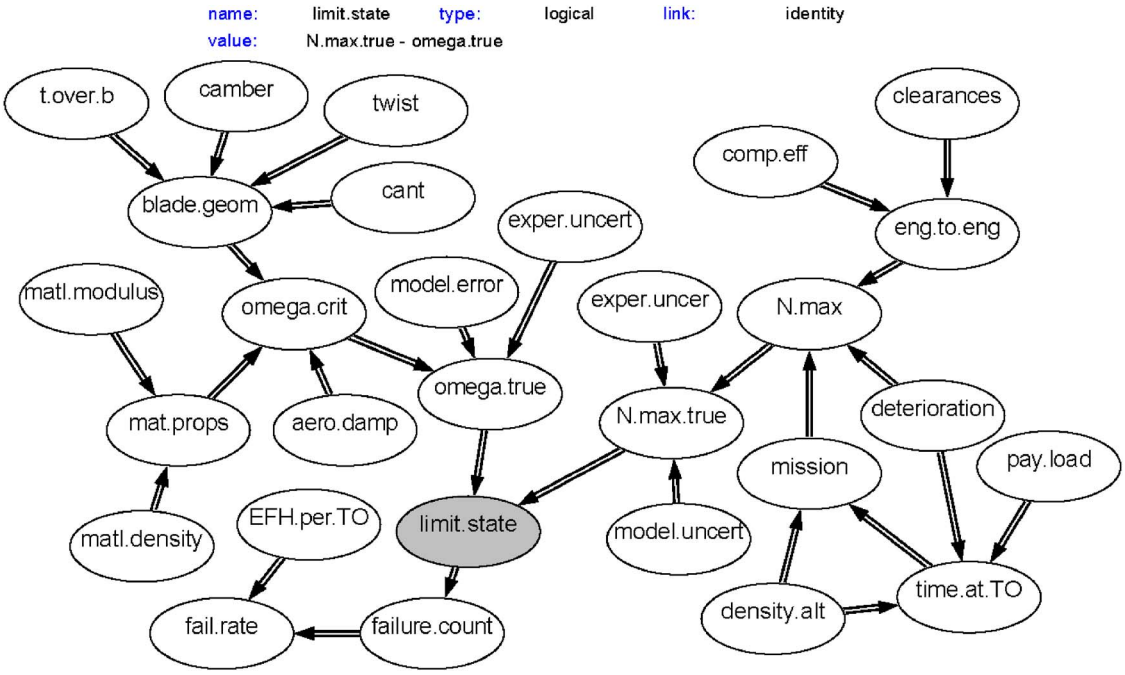


Fig. 1 WinBUGS DAG for probabilistic HCF

- *Production tolerances or variations in blade geometry* as they affect the fundamental mode natural frequencies.
- *Aerodynamic damping* is considered to be an interaction between vibration and aeroloading that is to be included for the case of this hypothetical design.
- *Material property statistics* based on standard acceptance testing results.
- *Measurement and modeling uncertainties* must be considered when converting the results of an FEA blade vibration model to predict the distribution of blade response for a fleet of engines. The simple hypothetical model used herein assumes that the FEA modeling error, and the experimental calibration uncertainties are independent probabilistic factors.
- *Engine-to-engine production variabilities* are considered in this hypothetical problem to stem from variations in tip clearances for the rotor and overall component efficiency for the rotor. The entire rotor would be considered here as it operates as a rigid ensemble regarding rotor speed variability. These would be aeroloading variabilities that are fully correlated across a given rotor.
- *Flight mission conditions* including atmospheric flight conditions represented by the density altitude for takeoff, the pilot profile that induces variability in the amount of time at TO power, and the flight manifest payload, are treated as fleet-wide sources of mission variability. Another mission variable is the *effective flight hours (EFH) per flight* required to convert the failure probability to a failure rate.
- *Deterioration* is taken to be an overall rotor/engine variable affecting the maximum rotor speed. Over time, deterioration will result in a higher HP rotor speed.
- *Measurement and modeling uncertainties* must also be considered when predicting the HP maximum rotor speed. Again, the three variables are linked through the model validation process that includes measurement and prediction uncertainties.

The probabilistic-HCF design is given as a Bayesian network in terms of the above RVs and their physical/logical connectivity in Fig. 1, a directed acyclic graph (DAG) based on the physics of the design problem. The graphic in Fig. 1 is taken directly from WINBUGS. Each of the logical nodes where RVs are combined is defined by algebraic functions supplied to the WINBUGS code. The outermost ovals are stochastic nodes where specific probability distributions are defined. These flow to logical nodes through double arrows leading through the logic of the defined physics to the outcome node, given by the name "limit.state."

Failure for this design problem occurs when the true rotor excitation frequency exceeds (a one-sided limit condition) the fundamental mode natural frequency (all in Hertz). Thus, the problem is a limit-state problem as represented by Eq. (2). In the more general case of probabilistic Campbell diagrams, failure might be taken as any excitation frequency within a defined interval containing the resonance frequency (two sided).

The outermost "parent" nodes in the DAG are stochastic nodes and represent the input data defined in terms of their mean values and precisions (inverse of variances). The statistics for each top-level, data input RV were selected for the example to make the example outcome seem reasonable. The limit.state node is the true engine-order frequency minus the true component resonance frequency. As can be seen, this is a typical *S-R* type of problem with the natural frequency node "omega.true" representing the stress variable and the maximum rotor speed node "N.max.true" representing the strength variable. The node "failure.count" simply records the number of limit-state excedances and normalizes this to the total number of simulation iterations for the MCMC chain.

The MCMC algorithm provides a Bayesian update probability value for each node (posterior) based on the values of all parent node prior values. In this way, all nodes are linked through Bayes' theorem to all other network nodes. The iterative process achieves a formal Markov chain linking priors to posteriors for each node of the network. The set of simulated outcome results is in the form of an empirical distribution of values of the limit-state node that is converted to a probability of failure. WINBUGS provides the mean and standard deviations along with histograms for each node.

The assumption in this example is that the input distributions for each of the RVs are completely specified. That is, there is no uncertainty due to limited data that would limit our confidence in

Table 1 WINBUGS output statistics

Node	Mean	SD	MC error	5.00%	Median	95.00%	Start	Sample
EFH.per.TO	2.5	0.1006	0.001408	2.334	2.499	2.666	1001	5000
N.max.true	11,160	1394	19.28	8879	11140	13500	1001	5000
aero.damp	0.5014	0.151	0.001946	0.2517	0.5009	0.7508	1001	5000
blade.geom	0.001368	0.0795	0.001092	−0.1302	0.001621	0.1318	1001	5000
camber	0.001167	0.1003	0.001438	−0.1662	5.42E-04	0.1655	1001	5000
cant	0.002301	0.1001	0.001218	−0.1599	0.001966	0.167	1001	5000
exper.uncert	$1.28\ 10^{-4}$	0.09929	0.001438	−0.1659	0.00206	0.1662	1001	5000
fail.rate	0.02365	0.09434	0.001317	0	0	0.3836	1001	5000
failure.count	0.0592	0.236	0.003269	0	0	1	1001	5000
mat.props	$9.22\ 10^{-5}$	0.02224	$3.00\ 10^{-4}$	−0.03576	−8.45E-05	0.0375	1001	5000
model.error	0.01502	0.008842	$1.26\ 10^{-4}$	0.001492	0.01497	0.02867	1001	5000
omega.crit	14,480	1190	15.32	12,500	14490	16410	1001	5000
omega.true	14,700	1885	29.34	11,660	14650	17860	1001	5000
t.over.b	0.001479	0.09849	0.001354	−0.1596	0.003146	0.1642	1001	5000
twist	−0.001282	0.1001	0.001364	−0.1624	−0.001852	0.1622	1001	5000

the input statistics. Thus, we have posed a standard probabilistic design problem with only a single probability of failure outcome.

A 5000-iteration set of statistics for the hypothetical probabilistic HCF problem follows in Table 1. The statistics are compiled for the 5000 samples taken after an initial burn-in run of 1000 iterations to assure a stable Markov chain. Table 1 provides the results for part of the network focusing on the stress side of the limit state, and it shows the empirical statistics derived for the forward MCMC simulations based on the input parameters for each independent RV, described by specified probability distribution parameters.

The standard deviation (SD) statistics for the nodes "failure.count" and "fail.rate" apply to the binary response of that variable and do not represent a lack of confidence in the failure rate statistic. The *mean* value for the node failure.count correspond to the actual number of times the limit state is violated while the node fail.rate automatically divides this number by the number of iterations.

For the real design problem, we are concerned with the case for which the input RV statistical parameters are not precisely known and we want to calculate the CI statistics for the probability of failure. In the interest of brevity, we leave it to the reader to accept the fact that if we create hyperdistributions for each of the input RVs, simulating in a Bayesian manner our uncertainty on the input variables, we would still get a single probability of failure statistic from the above example.

Bayesian Network Approach for a Confidence Interval on Failure Probability

A proposed alternative to the stated shortcoming of Bayesian network models in regard to simulating the inner- and outer-loop problems to get confidence interval results is based on the structure of Eq. (2). We approximate the conditional probability function in Eq. (1) by a response surface function of the conditional parameters, thereby algebraically representing the conditional dependence of the failure probability on the probability distribution parameters

$$\Pr_F(\alpha^*) \approx \Pr[\hat{\Pr}_F(\vec{\vartheta}, \vec{\varphi}) \leq \alpha^*] \tag{3}$$

The hat indicates a response surface approximation to the inner-loop integration for the probability of failure. We use a response surface (RS) approximation to this conditional probability that is generated by a design of experiments (DOE) approach in order to minimize the needed number of numerical experiments. A demonstration of the RS approach for the hypothetical probabilistic HCF problem follows.

Step 1. Generate the RS based on the inner-loop portion of the total probability integral, as indicated in Eq. (3). The statistical importance factors provide a selection basis for reducing the total set of numerical experiments to a more manageably sized set.

We used a simple approach for defining the statistical importance factors. We defined the factors to be the components of a vector in the RV space from the origin to the mean of the failure region (the mean of the values of the RVs for all simulations that incur failure). Figure 2 compares this simple approach using the WINBUGS results to the values obtained using a Nessus fast probability integration (FPI) routine. The agreement is quite good.

The RS for the hypothetical probabilistic HCF problem is thereby reduced to a function of four independent RVs with two independent probability distribution parameters per RV. The utilized RVs are the two experimental uncertainties, the thickness-to-chord geometric uncertainty, and the clearances uncertainty.

We used a slightly modified form of the same Bayesian network model to generate specific values of the conditional probability of failure in the usual DOE mode. The DAG we used is the same as the one for the HCF example given in Fig. 1 but with specified values (the DOE limits on each of the parameters) for each of the four most important RVs.

Step 2. Establish a DOE model for the system. Typically, this

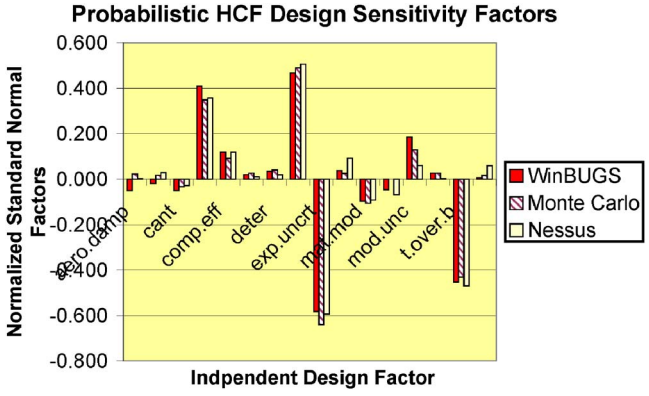


Fig. 2 Comparison of probabilistic sensitivity factors

Table 2 DOE parameter physical values

Parameter	Low	Mid	High
Mean clearance	−0.07	0.0	0.07
Mean experimental (two values)	−0.03	0.0	0.03
Mean geometry	−0.02	0.0	0.02
Standard-deviation clearance	0.07	0.105	0.18
Standard-deviation experimental (two values)	0.08	0.10	0.12
Standard-deviation geometry	0.09	0.105	0.12

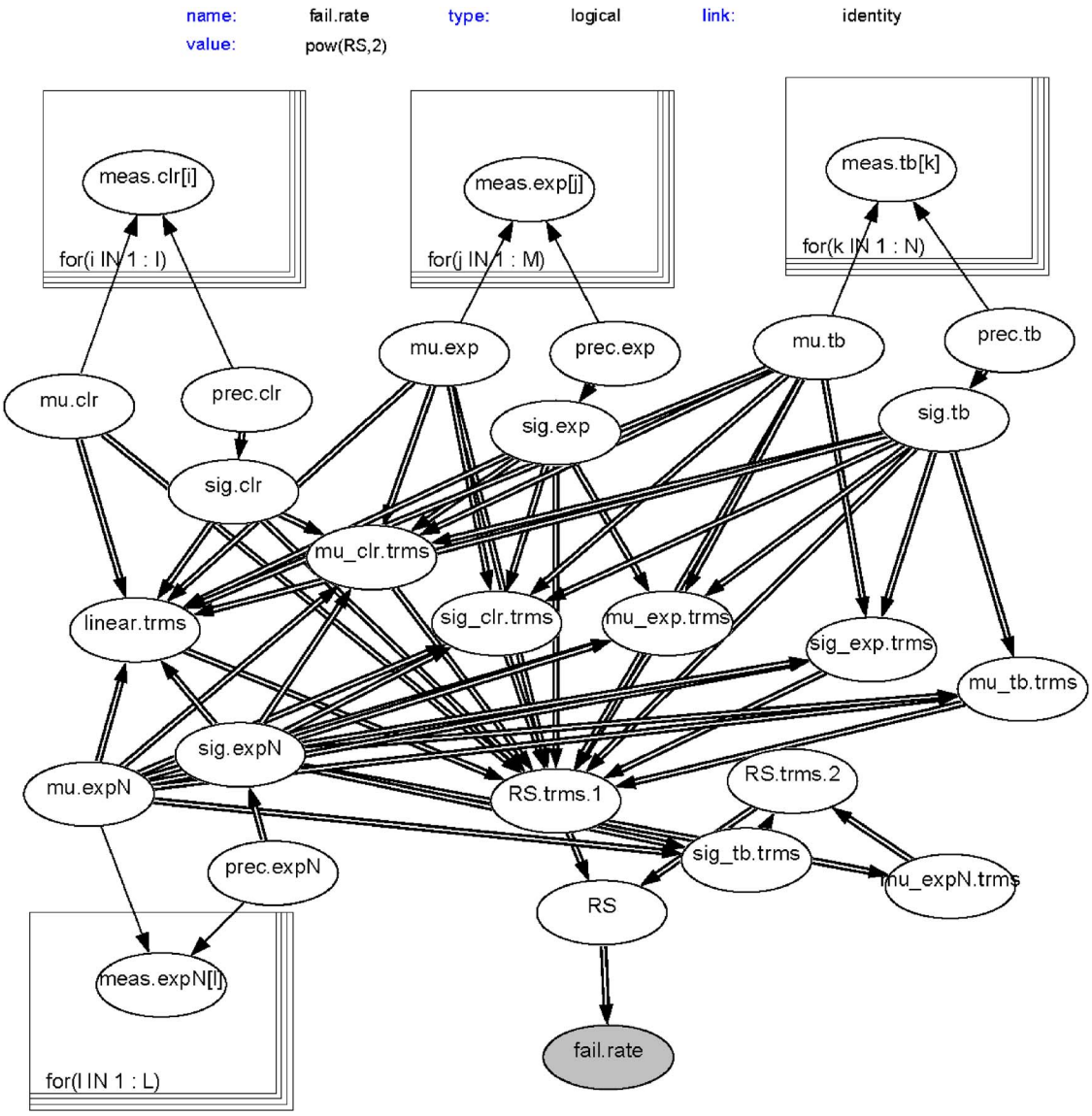


Fig. 3 DAG for the inner-loop response surface approximation

involves the selection of a DOE strategy and then the selection of the ranges (high/low) to be used for each data variable in the model (eight for this problem). There is no unique DOE strategy, and it is likely that no two persons will totally agree on the best one. The DESIGN-EASE® software product [3] was used to facilitate this DOE exercise.

The example problem began by the assumption that the RS fit could require up to a quadratic representation. The "D-Optimal" initial design of experiments was selected in the software to define the required sets of numerical experiments to be performed. The experimental design contained 55 experiments for the RS. Most of these are at corners of the design space (high/low parameter conditions).

Step 3. Define high/low ranges for the eight parameters. Each of the four RVs for the DOE happen to have been defined as normal probability distributions. Furthermore, since each of these four is implicitly tied to experimental data, standard statistical confidence interval estimates for small data sets were used to define conservative ranges of high/low for the mean and standard deviations of the four RVs. We used 90% CI values about the means for the high/low values based on the simulated set of experimental data, starting with an assumed number of 11 samples for each variable.

The intent in the high/low range setting is simply to span the range of values for each of the key parameters sufficiently widely that the RS provides reasonable accuracy for its conditional evaluation. More research is really needed to address the accuracy issues involved in this step of the process. Table 2 gives the values of the parameters used in the DOE computations.

Step 4. Perform the DOE set of numerical computations. The required 55 computations were performed as separate simulations. The results were then used by DESIGN-EASE to fit a RS. The code suggested that fitting the square root of the failure rate was favored and that the two-factor interaction (2FI) gave the best fit to the data. Although many statistical validity checks are available, this RS fit produced an R-square value of 0.999. This overall fit of the probability of failure response surface is not likely to be as accurate in the tail of the distribution associated with a high confidence limit. That issue is beyond the scope of the current note.

Step 5. Implement the RS. The RS model above is then to be implemented as a Bayesian network model for evaluating the outer loop integrations in Eq. (3). The RS model is a complicated nonlinear algebraic function of the eight variables from the above DOE. The algebra is too lengthy, and its form is not germane to our point. We have implemented the equation in a Bayesian network as shown in Fig. 3. The inner nodes collect the various

Table 3 WINBUGS confidence interval statistics

Node	Mean	SD	MC error	5.0%	Median	95.0%	Start	Sample
fail.rate	0.05252	0.02974	3.902×10^{-04}	0.01705	0.04627	0.1075	1001	5000
mu.clr	0.01262	0.05551	7.19×10^{-04}	−0.07733	0.01282	0.1045	1001	5000
mu.exp	0.02636	0.05681	7.93×10^{-04}	−0.06631	0.02512	0.1208	1001	5000
mu.expN	0.002561	0.05566	7.358×10^{-04}	−0.08862	0.003871	0.08971	1001	5000
mu.tb	0.005352	0.05817	8.126×10^{-04}	−0.09001	0.004709	0.1001	1001	5000
sig.clr	0.1814	0.04614	8.053×10^{-04}	0.1237	0.1733	0.2648	1001	5000
sig.exp	0.1844	0.04589	5.774×10^{-04}	0.1253	0.1759	0.270	1001	5000
sig.expN	0.1799	0.04485	7.214×10^{-04}	0.1232	0.1721	0.2591	1001	5000
sig.tb	0.186	0.04702	7.836×10^{-04}	0.1268	0.1777	0.2770	1001	5000

algebraic terms in a convenient manner.

In order to best illustrate the Bayesian CIs, we have assumed now that the key RVs are defined by experimental databases rather than hyperdistributions. The point of the illustrated confidence interval type of result will be the same either way, but this approach allows us to simulate the effect of varying sample sizes on the resulting CI for the probability of failure.

WINBUGS provides a graphical way to implement data sets into the Bayesian network. The WINBUGS data tablets shown in Fig. 3 represent each of the four key RVs. Each data table has its own integer range, the values of which are input by a separate data statement.

In this hypothetical simulation, we have chosen to simulate experimental data sets that are equal in size. By so doing, we can examine the convergence rate of the CI for the probability of failure as a function of a single dataset size parameter. A random-number generation algorithm was used to simulate the four independent sets of experimental data based on an underlying normal distribution.

A normal prior $N(0.0,10^{-6})$ (note that precision is used by WINBUGS as the second parameter for the normal distribution) was used for each mean. A Gamma prior $G(0.1,0.1)$ was used for the precision variable. The Bayesian network simulation statistics derived as posterior estimates for the case of 11 data points in each dataset are given in Table 3.

The probability distribution for the simulated failure rate conditionally dependent on the experimental data is shown graphically as a probability density function in Fig. 4. This WINBUGS simulated density function was validated using a direct Monte Carlo simulation of the problem. The 95% bound on the failure rate is 0.1075.

The mean or expected failure rate with perfectly defined input probability distribution parameters was computed in the previous section to be 0.024. The results are based on uncertain input probability distributions derived from the Bayesian analysis has a mean failure rate of 0.053. The failure rate now does have a probability distribution compiled by the program that is conditionally dependent on the parameters of the input variables. These parameters are described following the Bayesian network analysis as posterior distributions.

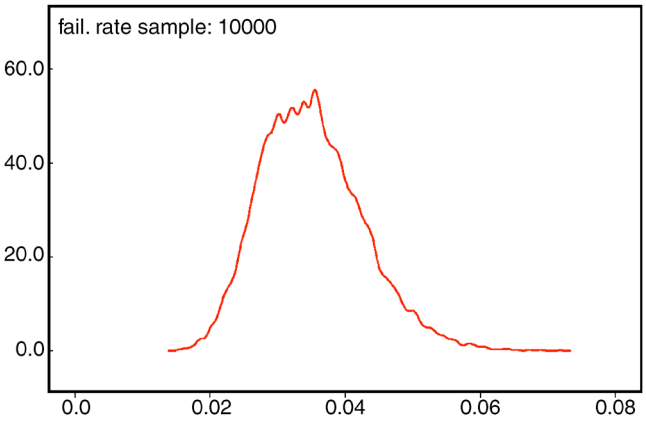


Fig. 4 Failure rate probability density function simulation results for sample size=11

Equal sample size sets of data were generated for $n=21,51$, and 101 samples. Table 4 summarizes the results for the simulations as a function of sample size.

The plot in Fig. 5 compares the MCMC-derived confidence interval to some analytical bounds. Of course, there are no theoretical bounds that truly apply to the systems problem formulation with joint probability distribution results taking multiple marginal distribution updates.

The failure simulation is a form of Bernoulli trial. For large numbers of trials, the expected probability of failure for a Bernoulli trial has an approximately normal distribution [4], pp. 393–396]. The variance for the large sample case can be written in the following approximate form:

$$\sigma^2_{\hat{\mathrm{Pr}}_F} = \frac{\hat{\mathrm{Pr}}_F(1-\hat{\mathrm{Pr}}_F)}{n} \tag{4}$$

although there is no theoretical justification for the use of this statistic in plotting bounds, we have plotted the results shown in Fig. 5 by replacing the probability of failure with the failure rate. The result is listed as "theory." The sample size is taken to be the data set sizes (11, 21, 51, 101) in spite of the fact they are not small. Nonetheless, the statistic in Eq. (4) is in reasonable agreement with the MCMC upper bound at the 95% confidence level set for both. The plot in Fig. 5 also includes confidence bounds computed by the FPI code [5].

Step 6. Evaluate the relative importance of the RS factors. The same approach for defining the statistical importance factors was used for the RS model. The simulation results for values of the failure rate greater than the 95% confidence bound value were aggregated such that the mean design condition for the points outside the 95% limit could be determined. The mean design conditions for these points were converted to standard normal form using the empirical statistics from the Bayes' updates. The results based on a 5000 iteration sample with 51 data points for each of the four RVs are plotted in Fig. 6.

The resulting statistical importance factors do show a relative ranking for each factor. It seems that none of the factors is really small enough to be disregarded in setting the 95% CI bound on

Table 4 Convergence statistics for simulated failure rate

Sample size	Mean failure rate	SD failure rate	95% bound failure rate
11	0.053	0.030	0.108
21	0.050	0.180	0.083
51	0.031	0.009	0.047
101	0.033	0.008	0.044

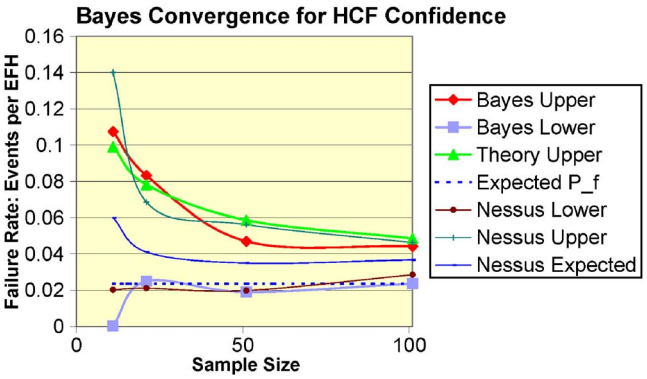


Fig. 5 Convergence of failure rate with sample size

the failure rate. The data do not show any obvious statistical trends that could be used to justify the ranking in importance of each of the statistical factors.

Conclusions and Recommendations

Bayesian network system models can be extended to include confidence intervals or failure rate bounds by adding a subsequent analysis step. A sensitivity analysis may be used to select those marginal probability distribution parameters that are the most important ones for the confidence interval estimation. A response surface representing the conditional probability of failure or failure-rate term can be constructed using a standard design of experiments approach for determining a response surface where each random parameter of importance is included.

Further work is needed to establish rational descriptions for use in modeling the uncertainties in the original random variable marginal probability distribution parameters. This work does not involve and new fundamental issues. The work should also address the most effective way to include the results of the validation studies for the physical models so that the proper probabilistic models for experimental and modeling errors are defined.

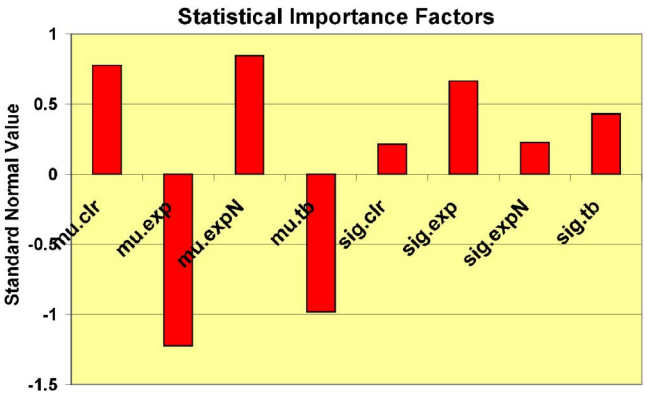


Fig. 6 Statistical importance factors for 95%-bound conditions

Acknowledgment

We gratefully acknowledge the many useful and critical discussions on the development of this note with Dr. Dan Ghiocel (STI, Inc.), Professor Sankaran Mahadevan (Vanderbilt University), Jim Griffiths (GEAE), and Chuck Annis (Statistical Engineering, Inc.). The primary author acknowledges the support of the AFRL Propulsion Directorate through a consulting subcontract to Universal Technologies Corporation, with Michele Puterbaugh as Contract Manager, to prepare this paper. The original work was done by the principal author as part of a DARPA "seedling" program under Dr. Leo Christodoulou funded through the AFRL Materials Directorate with Dr. James Larsen and Mr. Jay Jira as Program Managers.

Nomenclature

α = confidence interval (%)
F_X = cumulative distribution function for random variable, X
f_X = probability density function for random variable, X
$g_{X,Y}$ = joint probability density function
g_X = probability density function
μ_X = mean value of random variable X
Pr = probability function
$\Pr_F$ = probability of failure
$\hat{\Pr}_F$ = approximate probability of failure using response surface
Φ = standard normal variable [0,1]
$\vec{\varphi}$ = vector whose components are parameters of probability distributions
R = random variable
r = real values $-\infty \leq r \leq \infty$
S = random variable
s = real values $-\infty \leq s \leq \infty$
σ_X = standard deviation of random variable X
σ_i = standard deviation of ith random variable
$\sigma_T = \sqrt{\sum_{i=1}^{N} \sigma_i^2}$
$\vec{\vartheta}$ = vector whose components are parameters of probability distributions
X = random variable
x = real values $-\infty \leq x \leq \infty$
ξ = dummy random variable
Y = random variable
y = real values $-\infty \leq y \leq \infty$
Z = random variable
z = positive real values $0 \leq z \leq \infty$

References

[1] Spiegelhalter, D., Thomas, A., and Best, N., 2000, WinBUGS V. 1.3, User Manual, http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml
[2] Carlin, B. P., and Louis, T. A., 2000, *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd ed., Chapman & Hall/CRC, Boca Raton.
[3] Anon., *User Manual*, Stat-Ease, Inc., Hennepin Square, Minneapolis, www.statease.com
[4] Mood, A. M., Graybill, F., A., and Boes, D. C., 1963, *Introduction to the Theory of Statistics*, 3rd ed., McGraw-Hill, New York.
[5] Anon., 1998, *Nessus/FPI Theoretical Manual*," Version 2.4, Southwest Research Institute, San Antonio, pp. 41–44.

Vibration Characteristics of a 75 kW Turbo Machine With Air Foil Bearings

Kyeong-Su Kim

In Lee[1]
e-mail: inlee@asdl.kaist.ac.kr

Department of Aerospace Engineering,
Korea Advanced Institute of Science and Technology,
373-1 Guseong-dong, Yuseong-gu,
Daejeon, 305-701, Korea

Air foil bearings are very attractive bearing systems for turbomachinery because they have several advantages over conventional bearings in terms of oil-free environment, low power loss, long life, and no maintenance. However, most of the developed machines using air foil bearings are limited to small and high-speed rotors of 60,000–120,000 rpm, since the increase in power of turbomachinery requires lower rotor speed and greater loading in bearings, which makes it difficult to use air foil bearings for large machines. In this paper, a 75 kW turboblower using air foil bearings is introduced, and the vibration characteristics of the machine have been investigated experimentally under a wide range of operating conditions, including compressor surge in the performance test. The machine is designed to be fully air lubricated and air cooled, and its operating speed is 20,000–26,000 rpm with maximum pressure ratio of 1.8. The results show that the air foil bearings offer adequate damping to ensure dynamically stable operation in the whole range. [DOI: 10.1115/1.2718220]

Keywords: air foil bearings, turbo blower, environment-friendly system

Introduction

The air foil bearing is a noncontact bearing operated by the coupled interaction between hydrodynamic pressure of viscous fluid and elastic deformation of foil structures. It has several advantages over conventional bearings for applications to turbomachinery in terms of oil-free environment, high-speed operation, low power loss, low vibration, long life, and no maintenance. In addition, the machines using air foil bearings are much more reliable than those using conventional bearings because the structure of the rotor-bearing system is simple and an auxiliary system is not needed. Air foil bearings have been successfully used for many kinds of air cycle machines (ACMs), which are small, high-speed turbomachines [1]. Recently, studies on applications to extreme temperature machines, such as, cryogenic turboexpanders, turbochargers, and gas turbines, have actively progressed to extend the application boundary of air foil bearings [2–9]. And, advanced studies on the air foil thrust bearings, which may be an obstacle for application to turbomachines rather than the journal bearings with regard to the space required, load capacity, and heat generation, have been introduced [10,11].

Figure 1 shows the application spectrum of air foil bearings in terms of rotor speed and bearing shaft diameter [12]. Most of the developed machines using air foil bearings are limited to small and high-speed rotors of 60,000–120,000 rpm, although theoretically, air foil bearings are considered to be applicable to large rotors. Turbomachinery featuring high power and compact size has a characteristic that the increase in power requires lower rotor speed and greater loading in bearings, which makes it difficult to use air foil bearings for large machines. However, the industrial sector gradually demands oil-free turbomachines to be applied in the new fields in relation to the advantages of energy save, low vibration, low noise, and environment-friendly system. The turboblower is one of them. It is an industrial machine that supplies large amounts of air or gas flow with low-pressure ratio of 1.3–1.8. It is widely operated in the field of sewage treatment, the cement industry, ventilation systems, chemical process engineering, etc. An advanced turboblower uses a high-speed motor to drive an impeller assembled at the same rotor, and a fully air lubricated and air cooled system without any kind of liquids is required in order to realize a compact, highly reliable, and maintenance-free system. It is noteworthy that the turboblower is one of the most advanced applications of air foil bearings because rotor speed must be designed to be no more than 30,000 rpm for maximizing its total efficiency, and this makes the weight of rotor relatively high. An investigation on such applications is hardly available although remarkable achievements for the load capability of air foil bearings have been reported in experimental rig tests [13–15]. Swanson et al. [16] introduced the design and testing approach for a high-speed turbocompressor operating above the first bending critical speed with air foil bearings, which is expected to be one of the most advanced applications of air foil bearings with regard to rotordynamics. However, the machine falls within the purview of the applied region of air foil bearings in Fig. 1 since the operating speed of the machine is 70,000 rpm.

In this paper, a turboblower using air foil bearings is introduced. The vibration characteristics of the machine have been investigated experimentally under a wide range of operating conditions, including compressor surge in the performance test. And, the performance of air foil bearings, such as lift-off speed, damping coefficient, and stiffness, are estimated from coast-down data.

Turbo Blower Using Air Foil Bearings

A turboblower having a 75 kW high-speed motor and its rotor configuration are shown in Fig. 2. The machine has maximum pressure ratio of 1.8, and its operating speed range is 20,000–26,000 rpm. The rotor consists of a high-speed motor, a cooling fan, an impeller, and rpm wheel, where a motor element, hatched in Fig. 2, is assembled into the shaft with interference fit. The mass of rotor assembly is 11.5 kg, and the rotor is supported by bump-type [17] air foil bearings. Figure 3 shows the thrust bearing and front journal bearing employed, and the geometric parameters are described in Fig. 4 and Table 1. The foil structures were fabricated by plastic forming of spring-tempered sheet metal

[1]Author to whom correspondence should be addressed.

Submitted to the ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received March 10, 2006; final manuscript received July 15, 2006. Review conducted by Nagaraj Arakere.

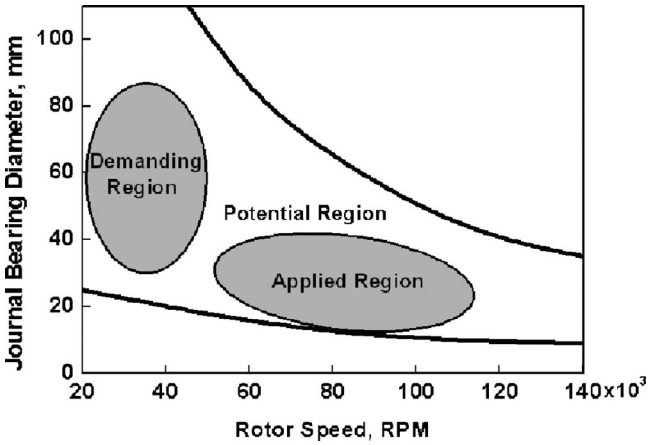
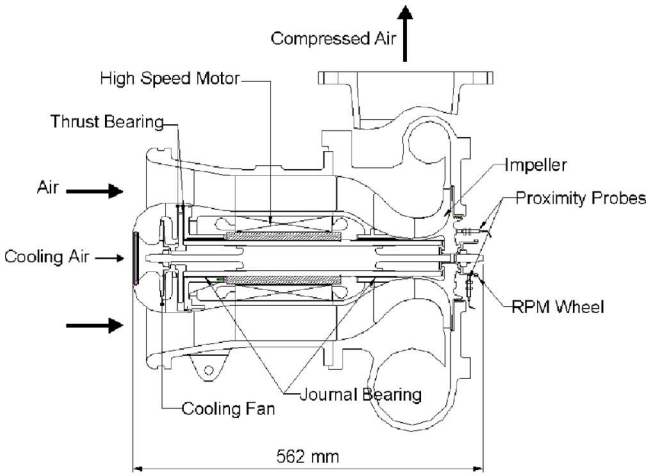


Fig. 1 Application spectrum of air foil journal bearings

using molds, and they were fixed by spot welding. Polytetrafluoroethylene coating was applied to the top foils and the facing surfaces of rotor were hardened above Hrc=50.

The machine is designed to be an environmental-friendly system that is fully air lubricated and air cooled. According to the preliminary analysis and the component's test results, it has been known that most of the heat is generated in the motor and the air foil thrust bearings, whereas air foil journal bearings produce much lower power loss, ~20% of the thrust bearings'. Therefore, the level of heat generation must be controlled within the capacity of the air-cooling mechanism provided by a cooling fan and compressor inlet flow. However, the cooling flow is very complicated and highly influenced by the operating conditions of the compressor, since the cooling flows for the bearing and motor affect each other and are eventually connected to the main airflow of the compressor. Therefore, the power-loss estimation is the most important design factor for an air cooled turbomachine because this parameter directly influences system performance and efficiency. Furthermore, a well-designed cooling scheme is essential to keep the temperatures of the bearing and the motor below their limit, and in this regard, it may be impossible to find a source of cooling



(*a*)

(*b*)

Fig. 2 Turbo blower and its rotor: (*a*) Layout of the turbo blower and (*b*) rotor assembly

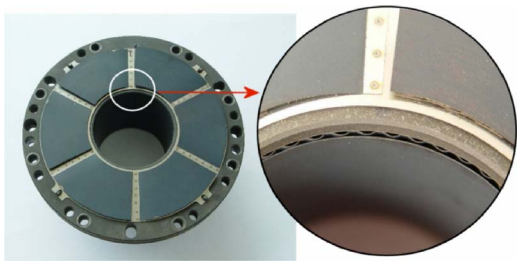
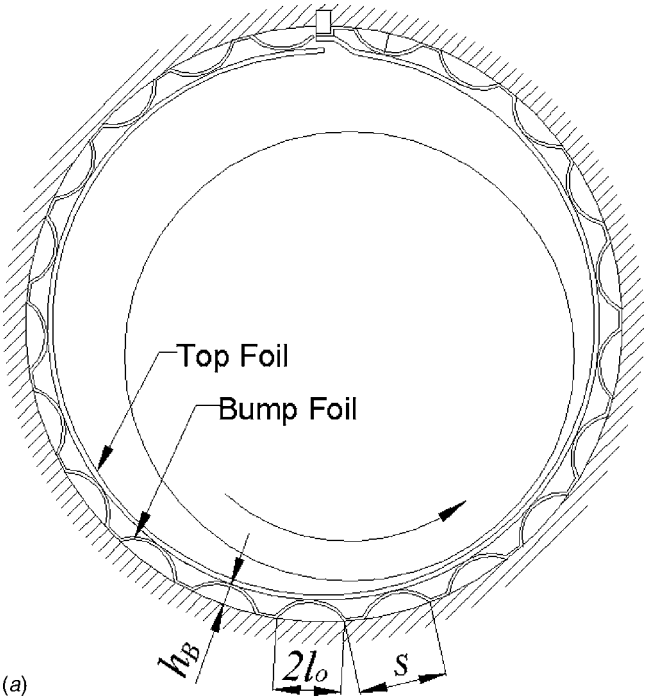
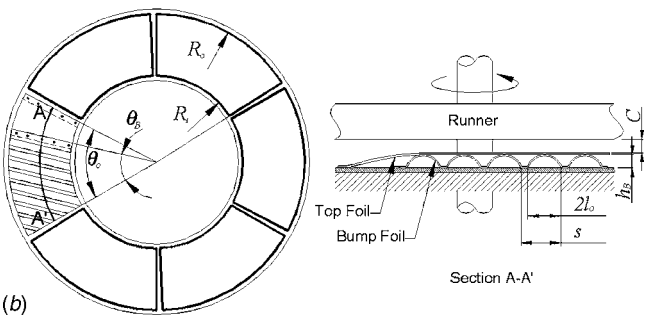


Fig. 3 Thrust bearing and front journal bearing



(*a*)



(*b*)

Fig. 4 Geometry of air foil bearings: (*a*) Journal bearing and (*b*) thrust bearing

Table 1 Parameters of air foil bearings

Journal Bearing	
Shaft diameter	60 mm
Axial length	60 mm
Nominal clearance	0.1 mm
Number of pad	1
Thrust Bearing	
Bearing inner radius	35 mm
Bearing outer radius	70 mm
Nominal clearance	0.2 mm
Number of pad	6
Bearing coating	polytetrafluoroethylene
Top foil thickness	0.2 mm
Top foil material	Inconel X750
Bump foil thickness	0.15 mm
Bump foil height	0.8 mm
Bump foil pitch	4.0 mm
Bump foil length	3.0 mm
Angular extent of bump foil	21.8 deg
Angular extent of pad	55.6 deg
Bump foil material	Inconel X750

flow when the power loss is excessive. Thus, air foil bearings must be optimally designed to achieve not only higher load capacity but also lower power loss.

The compressor performance is defined by a map, where the x-axis is airflow rate and the y-axis is discharge pressure. Compressor choking and surge are the boundaries of the operating region. The former presents the operating point with the maximum airflow rate, and the latter indicates the operating point with the maximum discharge pressure. When compressor surge occurs, a repetitive abnormal pressure distribution is generated in the compressor. This produces vibratory fluctuation of the thrust load, which can produce rotor instability and damage the bearings. Every turbomachine is designed and controlled to stay inside the boundaries of the operating region since compressor surge produces abnormal dynamic loads as noted above. Nevertheless, situations in actual industrial applications may arise where unanticipated high pressure is generated, resulting in surge. In this case, the bearings must have sufficient load capacity to withstand the hostile environment and protect the machine from failure.

Compressor Test Rig and Instrumentation

Figure 5 shows the compressor performance test rig and its instrumentation, where the prototype of the turboblower was directly employed to drive the compressor. A bell mouth was used to accurately measure the airflow rate into the inlet of the compressor at specified operating conditions—rpm and discharge pressure. Discharge pressure was controlled by a throttling valve located at discharge duct as shown in Fig. 5. Three eddy current-type proximity probes were used. Two of them were installed at the point of the rpm wheel in the x-y direction to measure rotor vibration, and the other one was installed at the point of impeller disk in the z direction to evaluate thermal growth and axial movement of the rotor. Pressure pulsation was monitored at the scroll, which is the compressor discharge section, using a dynamic pressure transducer (kulite ETM 375) that can effectively detect the compressor surge. The temperature of cooling air at the outlet section of the loaded front thrust bearing was measured to detect the rupture. Temperatures for the thrust bearing, motor stator, and discharge air were measured by using K-type thermocouples. The rotor speed was accurately measured using a magnetic pickup on the rpm wheel, which has 27 teeth. All the dynamic data have been recorded and analyzed utilizing a fast Fourier transform (FFT) analyzer.

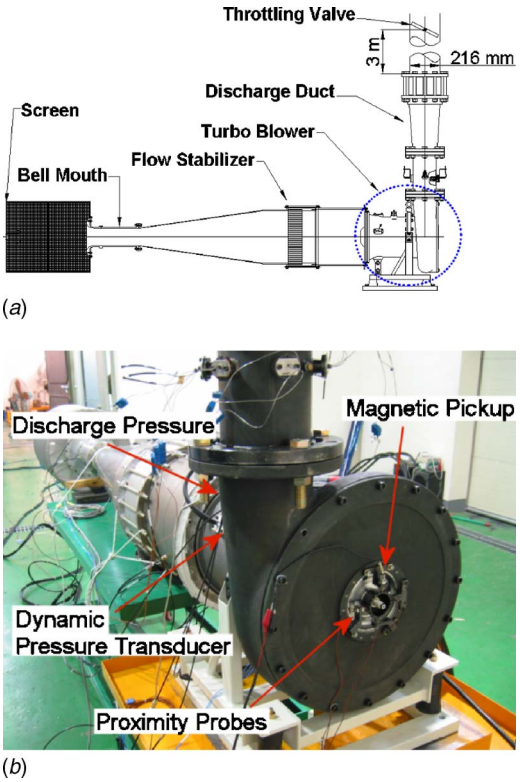


Fig. 5 Performance test rig and instrumentation: (*a*) Test rig and (*b*) instrumentation

Test Results and Discussion

Vibration Characteristics at Surge. Figures 6 and 7 shows the performance test results, where discharge pressure was increased step by step until surge occurred while rotor speed was fixed at 25,000 rpm. Each test step was maintained for enough time to get steady-state data. Upon the surge, discharge pressure was decreased as soon as possible by fully opening the throttling valve so as to escape the surge regime.

Temperatures and axial displacement were plotted in Fig. 7. It is noted that the machine has been heated up slowly in 1000 s until the motor temperature was maximized, and thereafter the motor temperature diminished as the discharge pressure increased since power consumed to drive the centrifugal compressor reduced accordingly. The axial displacement, which was measured relatively between the scroll case and the impeller disk, showed an opposite trend to the motor temperature. The axial displacement was 1.16 mm at the beginning, and then minimized by the

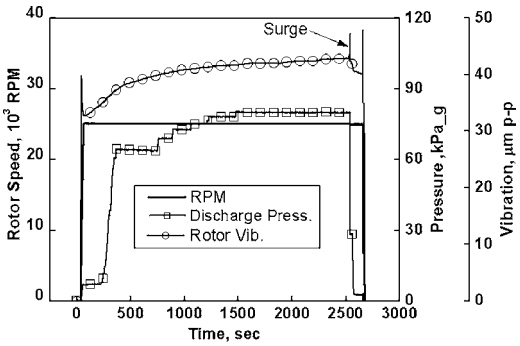


Fig. 6 Rotor speed, discharge pressure, and rotor vibration versus time

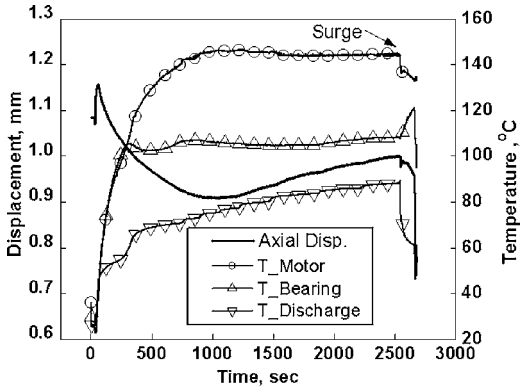


Fig. 7 Temperatures and axial displacement of rotor versus time

reduction of 0.25 mm, which means that the rotor assembly has been thermally expanded or grown as temperatures increase. The axial displacement was also influenced by the axial movement of rotor by the change of thrust load because it can be seen that the axial displacement ascended stepwise as discharge pressure went up. These features of the axial displacement are very important to evaluate the running tip clearance of an impeller, which affects the compressor efficiency. The thrust bearing temperature and the motor temperature were stabilized around 110°C and 145°C, respectively, indicating that the two components are the main source of heat generation. Interestingly, it was observed that bearing temperature rose rapidly when the throttling valve was fully open. This is due to the complicated characteristics of the cooling network, which is out of the scope of the current study.

Rotor vibration was monitored at the end of the rotor; thus, it is a measure at a distance from the rear journal bearing. It is noted that the displayed vibration value is overall peak-to-peak amplitude, which includes mechanical runout of the rpm wheel and thermal bending of the rotor assembly as well as the actual vibration. The measured vibration level increased as temperatures change by the thermal bending even though rotor speed remained constant. The thermal bending is attributed to the non-axisymmetric thermal expansion of the rotor due to the non-axisymmetricities of the interference fitting for the motor element, which is depicted in Fig. 2, manufacturing tolerances, and assembly tolerances. The amount of the thermal bending must be minimized and controlled within the load capacity of air foil bearings. It is shown that there were rises in vibration level during start-up and coast down of the machine due to the rigid-mode critical speeds of the rotor. The vibration level gradually increased from 33 μm to 43 μm as test conditions changes at 25,000 rpm, and it was observed that there was an abrupt increase up to 47 μm in vibration amplitude at surge. Generally, it is known that compressor surge generates low-frequency sound and fluctuations in inlet and discharge pressure. This fluctuation produces a related change in the rotor thrust load that may be detrimental to the air foil thrust bearing. Additionally, it excites the rotor-bearing system, resulting in vibration problems in terms of rotor dynamics. The most effective method in monitoring surge is to install a pressure transducer directly on the airflow channel and measure dynamic pressure. Figure 8 presents the time-domain signal of the pressure pulsation at the very moment of entering surge and its spectrum at surge. It is seen that the peak-to-peak amplitude of the pressure pulsation is ~14% of the static pressure level and its main frequency is 7.5 Hz. There are also small peaks related to the harmonics of the main frequency.

Figures 9 and 10 show the rotor orbits and their frequency spectrums at surge, and just after surge, respectively. From the plots, it is clearly indicated that rotor vibration is highly influenced by the surge and there are vibration components at low

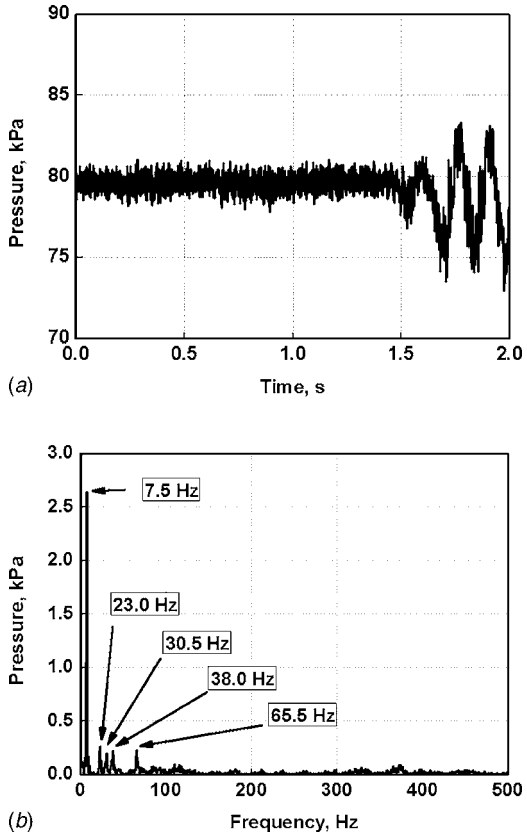


Fig. 8 Pressure pulsation measured at surge: (*a*) Time signal and (*b*) frequency spectrum

frequencies with significant level besides the synchronous rotor vibration of 417.5 Hz during the surge. The frequencies of 7.5 Hz and 15.5 Hz—this is nearly twice of 7.5 Hz—are related to the pressure pulsation caused by the surge, and there are small sidebands with the frequency difference of 7.5 Hz around synchronous frequency. At the same time, the rotor orbit becomes swayed and unstable. These vibrations might damage not only the thrust bearing but also the journal bearings. However, the rotor orbit was recovered after escaping the surge, and the test was successfully completed without any problems. Figure 11 shows the frequency spectrum of the axial movement of rotor at surge, which has very similar characteristics to the rotor vibration. After the performance test, the machine was completely disassembled and a tear-down inspection revealed no damage to the bearings.

Characteristics of Air Foil Bearings. Figure 12 presents a waterfall diagram of rotor vibration in the vertical direction during coastdown from 25,000 rpm. It is seen from Fig. 12 that the rotor-bearing system is stable over a range of operating speeds because the major vibration component is synchronous frequency with negligible harmonics. The horizontal motions are very similar to the vertical ones, and it suggests that the air foil bearings are nearly isotropic.

The rotor speed decreases dramatically by the effect of aerodynamic drag at high speed, and the slowdown ratio is declined as rotor speed goes down until the shaft contacts air foil bearings. Rotor speed and synchronous rotor vibration during cost-down are displayed in Fig. 13. The liftoff speed of the air foil bearings can be estimated to be ~3000 rpm from the rotor-speed curve. Once the shaft touches the bearing, the rotor stops in a short time due to the friction between shaft surface and the coating of bearing.

Figure 13 has been transformed into the form of Bode plot in Fig. 14, where the synchronous rotor vibration is displayed with respect to the rotor speed. The vibration level is in zero-to-peak

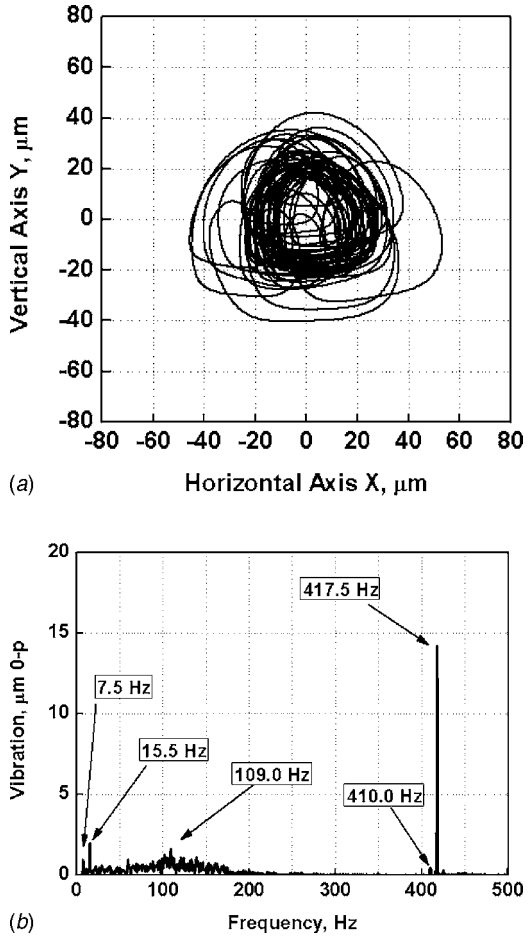


Fig. 9 Rotor vibration measured at surge: (*a*) Rotor orbit and (*b*) frequency spectrum

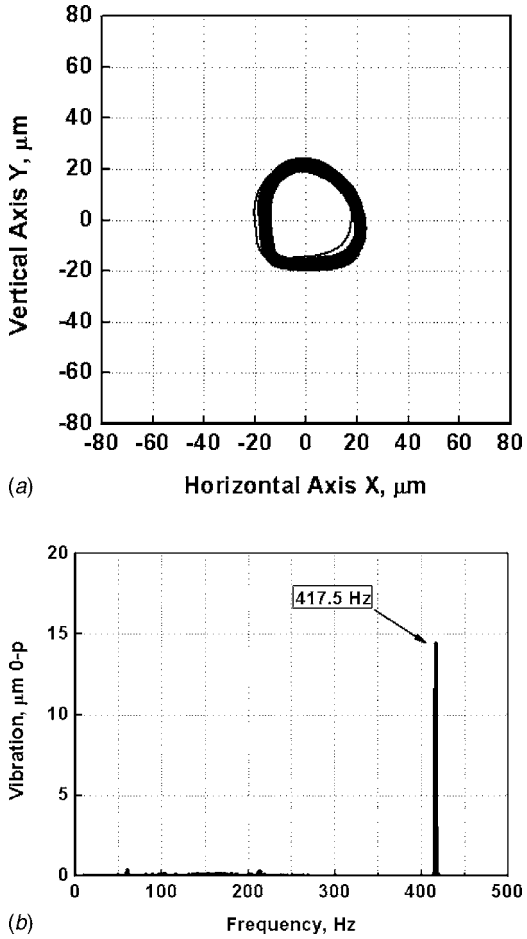


Fig. 10 Rotor vibration after being recovered from surge: (*a*) Rotor orbit and (*b*) frequency spectrum

amplitude, and a smooth curve using cubic spline connections was generated with the measured data indicated by symbols. It is shown in Fig. 14 that there is a critical speed at 9500 rpm, which is the most dominant rotordynamic response under the current unbalance loading condition. The mechanical runout of the rpm wheel with respect to the bearings must be compensated to extract the actual vibration amplitude in the plot. The mechanical runout, which is equal to the measured amplitude when the rotor is about to stop, is ~11.3 μm, and the maximum vibration amplitude A is 7.4 μm. The system damping with regard to the critical speed can be evaluated from this plot using an amplification factor [18]. Half-power points are f_1=8200 rpm and f_2=13,950 rpm with the critical speed of f_n=9500 rpm. Then, the amplification factor and the damping ratio can be calculated as follows:

$$\mathrm{AF} = \frac{f_n}{f_2 - f_1} = \frac{1}{2(c/c_c)} \tag{1}$$

The damping ratio c/c_c is 0.3026, and the critical damping c_c of the machine is 22,881 Ns/m. The system damping c is estimated to be 6924 Ns/m from these results, where most of the damping is expected to be provided by the air foil bearings. This result shows that the air foil bearings have sufficient damping to smoothly pass the critical speeds. The critical speeds of the turboblower were calculated varying support's stiffness as shown in Fig. 15, where each journal bearing was modeled as two linear springs located at the axial ends of the bearing. The first two critical speeds, which have rigid mode shapes, stand close to each other, and the third one, with a bending mode shape, is far beyond the operating range. The first two critical speeds are not differentiated clearly in Fig. 14 since the system damping is big enough to combine them into a single response. The stiffness of the journal bearing is approximated to be 2.0×10^6 N/m–3.0×10^6 N/m from Figs. 14 and 15.

Durability of Air Foil Bearings. The main objectives of the rig test were to evaluate compressor performance and identify the compressor surge line not to overload air foil bearings. Therefore, the surge intensity was controlled carefully to minimize damage of the hardware. Consequently, the employed air foil bearings left no traces of damage by inspection after the aforementioned surge

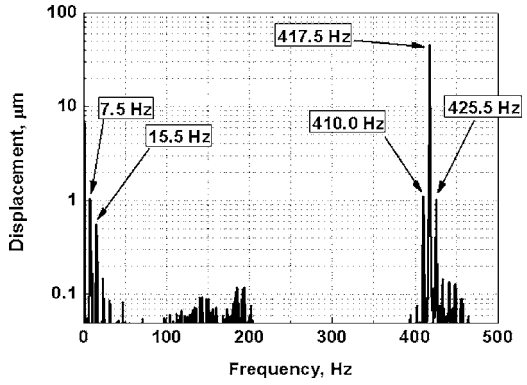


Fig. 11 Frequency spectrum of axial movement of rotor at surge

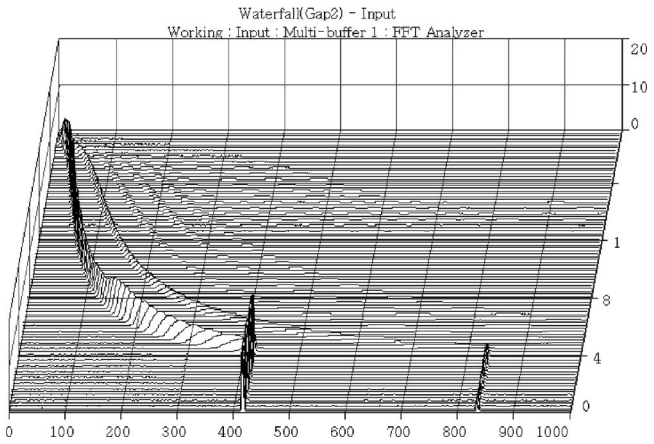


Fig. 12 Vertical waterfall diagram during coast down

test. It says that air foil bearings are robust enough to cope with a mild aerodynamic instability for the industrial application. However, a deep surge may be fatal to the bearings. Figure 16 shows an example of bearing damage after a deep surge in the field, where excessive aerodynamic loads and vibrations left hammered marks on the top foils of both journal bearing and thrust bearing. The top foils were reusable; however, the bump foils were permanently deformed and replaced. Figure 17 presents other pictures of air foil bearings after one year of operation in the field, and the bearings were reassembled for operation. It is clear that air foil bearings work in spite of local wearing out of the coating on the top foils unless there is plastic deformation.

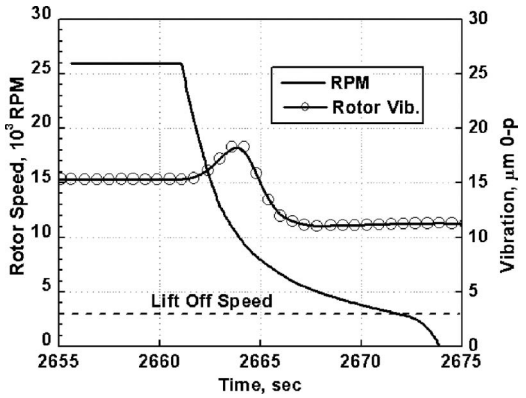


Fig. 13 Rotor speed and synchronous rotor vibration during cast down

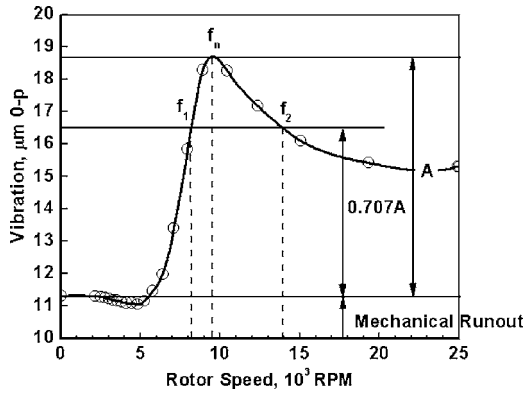


Fig. 14 Synchronous rotor vibration versus rotor speed during coast down

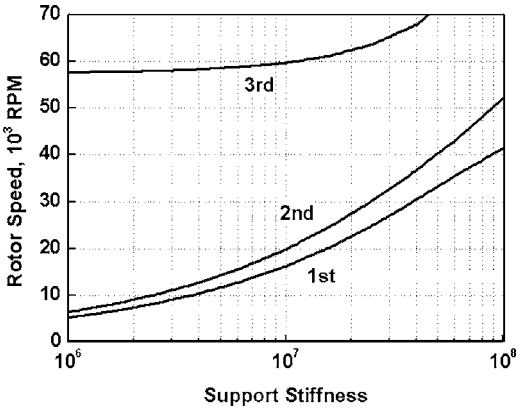


Fig. 15 Critical speeds versus support stiffness

Conclusion

It is summarized from the current investigation that air foil bearings offer excellent performance and reliability for application to a 75 kW turboblower, even at compressor surge, and they provide sufficient damping to smoothly pass critical speeds ensuring dynamically stable operation in the whole range. It is also shown that the characteristics of the air foil bearings, such as liftoff speed, damping coefficient, and stiffness, can be estimated from experimental data. Conclusively, this work has demonstrated that air foil bearings have potentialities in the development of various competitive oil-free turbomachinery, which has relatively low operating speed in industrial fields.

Acknowledgment

The work reported herein was supported by Neuros Co., Ltd. The authors express their appreciation to the company for their help in preparing the test rig.

Nomenclature

AF = amplification factor
C = nominal clearance (mm)
c = damping (Ns/m)
$c_c=2m\omega_n$ = critical damping (Ns/m)
f_n = critical speed (rpm)
h_B = height of bump foil (mm)
$2l_o$ = length of bump foil (mm)
m = rotor mass (kg)

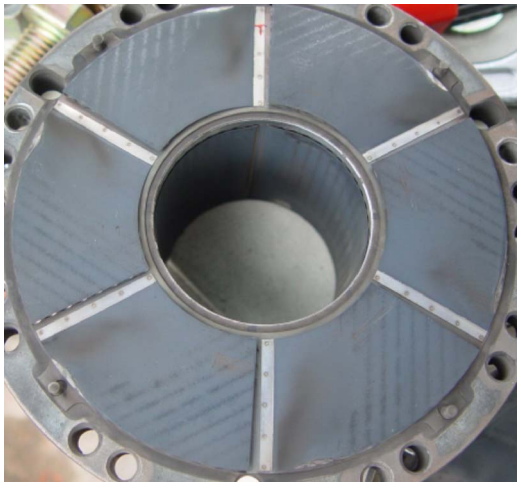
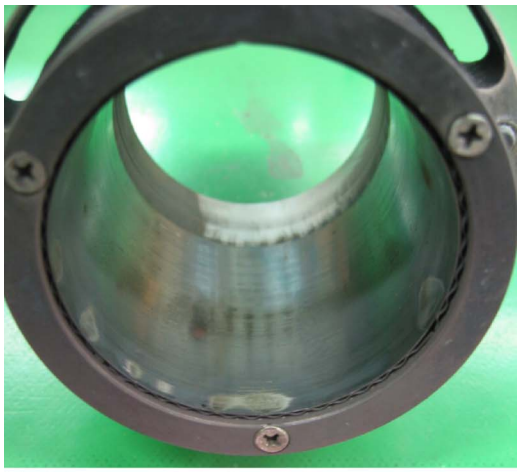


Fig. 16 Bearing wear pattern after a deep surge in the field



(*a*)

(*b*)

Fig. 17 View of air foil bearings after one year operation in the field: (*a*) Rear journal bearing and (*b*) thrust bearing

R_i = thrust bearing inner radius (mm)
R_o = thrust bearing outer radius (mm)
s = pitch of bump foil (mm)
θ_B = angular extent of bump foil (degree)
θ_O = angular extent of pad (deg)
$\omega_n = 2\pi f_n$ = natural frequency (rad/s)

References

[1] Agrawal, G. L., 1997, "Foil Air/Gas Bearing Technology—An Overview," ASME Paper No. 97-GT-347.
[2] Andres, L. S., 1995, "Turbulent Flow Foil Bearings for Cryogenic Applications," ASME J. Tribol., **117**, pp. 185–195.
[3] Yao, H., Quan, H.-Y., and Chen, C.-Z., 1997, "Development of Foil Journal Bearing for Small High Speed Cryogenic Turboexpander," International Cryogenic Engineering Conference, 1, pp. 233–238.
[4] DellaCorte, C., Fellenstein, J. A., and Benoy, P. A., 1999, "Evaluation of Advanced Solid Lubricant Coatings for Foil Air Bearings Operation at 25° and 500°C," Tribol. Trans., **42**(2), pp. 338–342.
[5] Heshmat, H., and Walton, II, J., 2000, "Oil-Free Turbocharger Demonstration Paves Way to Gas Turbine Engine Applications," ASME Paper No. 2000-GT-620.
[6] Kirshmann, A. E., and Agrawal, G. L., 1998, "High Temperature Foil Air Bearing Development for A Missile/UAV Engine Application," JANNAF Propulsion Meeting, No. 675, pp. 101–108.
[7] Shimizu, N., Matsushima, M., Ishii, I., and Iwadate, H., 2003, "Research of High Speed Shaft System Using Air Foil Bearing," HONDA R and D Technical Review, **15**(2), pp 93–100.
[8] Chen, S. H., and Arora, G. K., 1986, "Development of a 1000F, 10,000 Start/Stop Cycle Foil Journal Bearing for A Gas Turbine Engine," AIAA/ASME/SAE/ASEE 22nd Joint Propulsion Conference, Paper No. 86-1457, A87-23253.
[9] Swanson, E. E., Walton, II, J. F., and Heshmat, H., 2002, "A Test Stand for Dynamic Characterization of Oil-Free Bearings for Modern Gas Turbine Engines," ASME Paper No. GT2002-30005.
[10] Heshmat, C. A., Xu, D., and Heshmat, H., 2000, "Analysis of Gas Lubricated Foil Thrust Bearings Using Coupled Finite Element and Finite Difference Methods," ASME J. Tribol., **122**(1), pp. 199–204.
[11] Bauman, S., 2005, "An Oil Free Thrust Foil Bearing Facility Design, Calibration, and Operation," NASA/TM-2005-213568.
[12] Heshmat, H., 2000, "Operation of Foil Bearings Beyond the Bending Critical Mode," ASME J. Tribol., **122**(2), pp. 192–198.
[13] Heshmat, H., 1994, "Advancements in the Performance of Aerodynamic Foil Journal Bearings: High Speed and Load Capability," ASME J. Tribol., **116**(2), pp. 287–295.
[14] Swanson, E. E., and Heshmat, H., 2000, "Capabilities of Large Foil Bearings," ASME Paper No. 2000-GT-620.
[15] Swanson, E. E., Heshmat, H., and Walton, II, J., 2002, "Performance of a Foil-Magnetic Hybrid Bearing," ASME J. Eng. Gas Turbines Power, **124**, pp. 375–382.
[16] Swanson, E. E., Heshmat, H., and Shin, J. S., 2002, "The Role of High Performance Foil Bearings in an Advanced, Oil-Free, Integral Permanent Magnet Motor Driven, High-Speed Turbo-Compressor Operating Above the First Bending Critical Speed," ASME Paper No. GT2002-30579.
[17] DellaCorte, C., and Valco, M. J., 2000, "Load Capacity Estimation of Foil Air Journal Bearings for Oil-Free Turbomachinery Application," NASA/TM-2000-209782, NASA, Cleveland.
[18] Ehrich, F. F., 1992, *Handbook of Rotordynamics*, McGraw-Hill, New York, p. 4.28.

Rotordynamic Performance of a Rotor Supported on Bump Type Foil Gas Bearings: Experiments and Predictions

Luis San Andrés
Mechanical Engineering Department,
Texas A&M University,
College Station, TX 77843-3123

Dario Rubio
Bechtel Corporation,
3000 Post Oak Blvd.,
Houston, TX 77056

Tae Ho Kim
Mechanical Engineering Department,
Texas A&M University,
College Station, TX 77843-3123

Gas foil bearings (GFBs) satisfy the requirements for oil-free turbomachinery, i.e., simple construction and ensuring low drag friction and reliable high speed operation. However, GFBs have a limited load capacity and minimal damping, as well as frequency and amplitude dependent stiffness and damping characteristics. This paper provides experimental results of the rotordynamic performance of a small rotor supported on two bump-type GFBs of length and diameter equal to 38.10 mm*. Coast down rotor responses from* 25 krpm *to rest are recorded for various imbalance conditions and increasing air feed pressures. The peak amplitudes of rotor synchronous motion at the system critical speed are not proportional to the imbalance introduced. Furthermore, for the largest imbalance, the test system shows subsynchronous motions from* 20.5 krpm *to* 15 krpm *with a whirl frequency at* ~50% *of shaft speed. Rotor imbalance exacerbates the severity of subsynchronous motions, thus denoting a forced nonlinearity in the GFBs. The rotor dynamic analysis with calculated GFB force coefficients predicts a critical speed at* 8.5 krpm*, as in the experiments; and importantly enough, unstable operation in the same speed range as the test results for the largest imbalance. Predicted imbalance responses do not agree with the rotor measurements while crossing the critical speed, except for the lowest imbalance case. Gas pressurization through the bearings' side ameliorates rotor subsynchronous motions and reduces the peak amplitudes at the critical speed. Posttest inspection reveal wear spots on the top foils and rotor surface.*
[DOI: 10.1115/1.2718233]

Introduction

High performance oil-free turbomachinery implements gas foil bearings (GFBs) to improve mechanical efficiency in compact units. GFBs fulfill most of the requirements of novel oil-free turbomachinery by increasing tenfold their reliability in comparison to rolling elements bearings [1]. Foil bearings comprise one or more arcuate foils supported by compliant structures distributed around the bearing surface. The compliant underlying structure, providing structural stiffness, comes in several configurations such as bump-type (see Fig. 1), leaf-type, and tape-type, among others. GFBs operate with nominal film thicknesses larger than those found in a geometrically identical rigid surface bearing, for example, since the hydrodynamic film pressure generated by rotor spinning "pushes away" the top foil and its supporting structure [2,3].

GFBs enable high speed operation and larger load capacity than rigid gas bearings [4]; in particular a tunable structural stiffness is characteristic in the latest configurations [5–7]. In GFBs, Coulomb-type damping arises due to the relative motion between the bumps and the top foil, and between the bumps and the bearing support wall [8,9].

For over three decades, air cycle machines for aircraft cabin pressurization use simple GFBs with conventional polymer coatings deposited on the top foil. Presently, commercial portable power (30 kW) gas turbine engines implement advanced GFBs to deliver oil-free units with the highest reliability. Oil-free turbomachinery moves steadily toward applications in turbochargers, small gas turbine engines, and auxiliary cryogenic turbopumps where savings in weight and complexity as well as maintenance reduction are most desirable [10,11]. Remarkable improvements in high temperature limits are obtained by using coatings (solid lubricants) on the top foils. Chen et al. [12] detail the manufacturing of bump-type foil bearings and present simple design guidelines along with an application to a cryogenic turbo expander.

Reference [12] provides a straightforward assessment of gas foil bearing technology. Yet, in spite of the progress advanced in recent years, foil bearing design is still largely empirical, each foil bearing being a custom piece of hardware, with resulting variability even in identical units, and limited scalability.

This paper presents measurements of the rotordynamic performance of a light-weight rotor supported on two bump-type GFBs and comparisons to predictions obtained from a rotordynamics model integrating linearized force coefficients for the gas film bearings. The test results also serve as benchmark for the calibration of predictive tools.

Description of Test Foil Bearing and Experimental Rotor/Bearing Facility

Figure 1 shows a schematic view of the test foil bearing, consisting of five arcuate bump strips, each with five bumps. The end of a strip is welded to the bearing sleeve while the other end is free. The top foil, coated with a spray-on Teflon® type coating of thickness 25.4 μm, consists of a thin metal sheet welded at the bearing sleeve at one end (spot weld) and free at the other end. In Fig. 1, the orientation of the top foil spot-weld with respect to the vertical (gravity) plane is noted.

Figure 2 shows the test rig for rotordynamic experiments of a hollow rotor supported on two bump-type foil bearings. A steel

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received June 3, 2006; final manuscript received December 14, 2006. Review conducted by Mihai Arghir. Paper presented at the ASME Turbo Expo 2006: Land, Sea and Air (GT2006), May 8–11, 2006, Barcelona, Spain. Paper No. GT2006-91238.

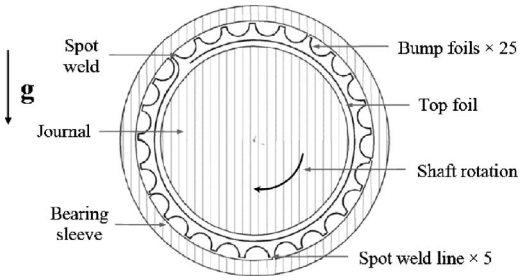


Fig. 1 Schematic view of a bump-type GFB. Location of top foil spot-weld relative to vertical plane as in tests.

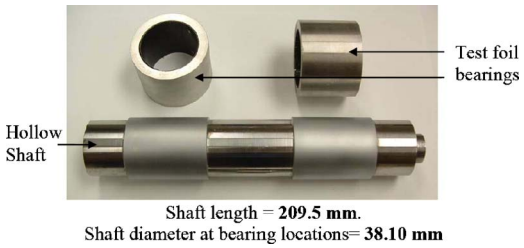


Fig. 3 Test rotor and foil bearings for rotordynamic tests

housing holds the test foil bearings in place and contains an internal duct to supply air pressure up to 690 kPa (100 psig) for cooling the bearings while in operation, if needed. The pressurized air impinges at the rotor midspan and flows axially through the test bearings. An electromagnet (EM) actuator, affixed to a groove in the housing, applies loads without contact at the rotor midplane. Typical air gaps between EM tip and the test rotor range from 0.25 mm to 0.50 mm. The EM consists of a slender rod made up of a high magnetic permeability material. Variable noncontacting loads are produced by sending dc current through the copper coil that surrounds the magnetic core. A strain gauge load cell, attached to the electromagnet, measures the reacting force acting on the test rotor.

Figure 3 shows a picture of the test rotor and foil bearings. Table 1 presents a summary of the bearings and rotor physical properties and geometry. The test rotor, made of steel AISI 4140, consists of a hollow shaft of length 210 mm and 38.10 mm in diameter at the bearing locations. A thin dense chrome coating, of thickness 25.4 μm, is applied to the rotor surface at the bearing locations to reduce friction and wear at the rotor/foil interface. Both rotor drive and free ends have 8 threaded holes at a 15.11 mm radius where imbalance masses are inserted.

An ac motor, 1.49 kW (2.0 HP), drives the test rotor up to a top speed of 25 krpm. The motor/rotor connection is through a miniature flexible coupling and a connecting shaft, see Fig. 2. The connecting shaft, made of steel AISI 4140, is 50.8 mm in length and 5.08 mm in diameter. The flexible coupling aided in isolating the drive motor dynamics from the rotor/bearing system; albeit making difficult the alignment of both components since the structural stiffnesses from the foil bearings are similar in magnitude as the coupling lateral stiffness.

Two pairs of eddy current sensors located at both rotor ends record the lateral rotor motions along the horizontal and vertical planes. Vibration signals from the eddy current sensors connect to a signal conditioner to bias the dc offset levels and then into a commercial data acquisition system for industrial machinery monitoring and diagnostic. A two-channel dynamic signal analyzer displays the frequency content of selected motion signals, and analog oscilloscopes display the unfiltered rotor orbits in real time.

While operating the test rig, the temperatures on the outer surface of the test foil bearings are monitored, as shown in Fig. 2. Also, as a result of the continuous current through the EM, heat is generated and temperature on the electromagnet surface rises. A controllable cooling system with oil flowing through copper tubes wound over the EM dissipates the generated heat.

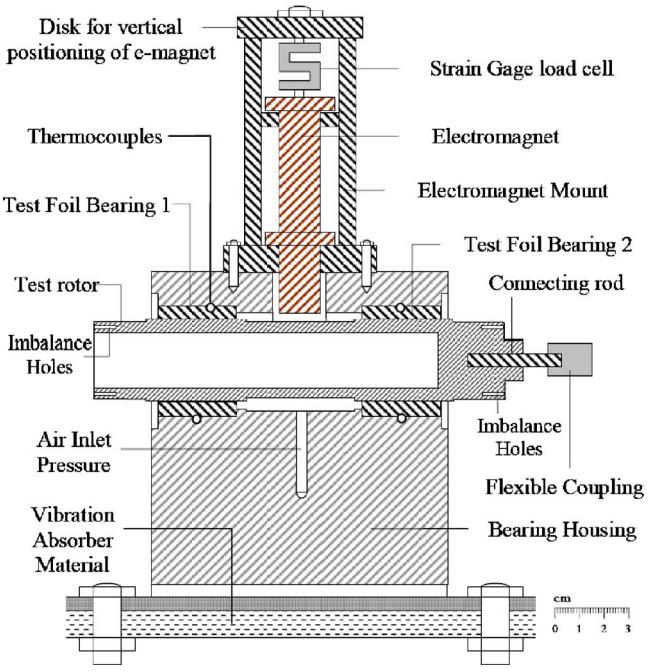


Fig. 2 Test rig for rotordynamic experiments of rotor supported on GFBs

Table 1 Summary of gas foil bearings and rotor physical properties and geometry[a]

GFB parameters: 25 bumps	SI units
Bearing inner diameter, $D=2R$	38.10 mm
Axial length, L	38.10 mm
Nominal clearance, c	35.5 μm
Bump pitch, p	4.572 mm
Bump length, l_0	4.064 mm
Foil thickness, t	0.102 mm
Bump height, h	0.381 mm
Poisson's ratio, υ	0.29
Bump modulus of elasticity, E	213 GPa
Free-free end bump stiffness, K_F	0.526 MN/m
Free-fixed bump stiffness K_W	0.876 MN/m
Rotor parameters	
Rotor modulus of elasticity E_R	193 GPa
Material density, ρ	7830 kg/m^3
Total mass, M	0.98 kg
Diameter at the bearing locations, D_j with thin chrome coating	38.20 mm
Total length, L_T	209.55 mm
Distance between bearing locations [6], L_S	100.58 mm
Distance between rotor CG to free end, x_G	125.73 mm
Transverse moment of inertia, I_T	37.1 kg cm^2
Polar moment of inertia, I_P	2.24 kg cm^2
Coupling lateral stiffness	1.63×10^5 N/m

[a]The uncertainties on physical dimensions are within 3%. Bump stiffnesses estimated using formula given in Ref. [7].

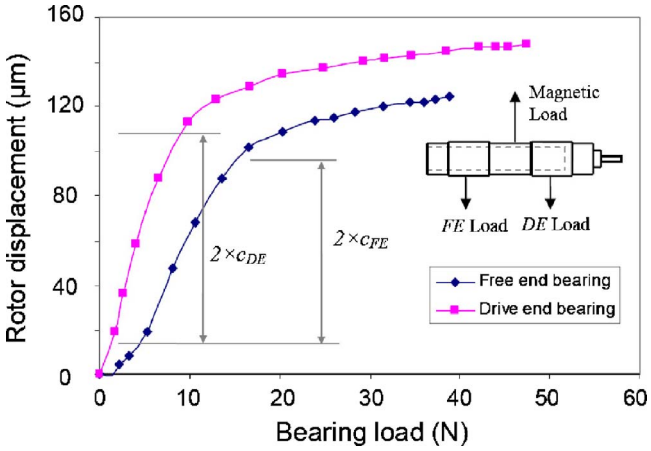


Fig. 4 Rotor static displacements at the free end and drive end bearings for increasing applied electromagnetic loads

Estimation of Clearance in Foil Bearings

The nominal clearance in a foil bearing is an ambiguous concept, largely unknown in most applications. The nominal clearance herein is defined as the air gap between the journal and the foil bearing nondeflected inner bore. A reliable estimation of radial clearance and static load of the test bearings is paramount for predictions of the foil bearing force performance.

Diametric clearances of the foil bearings are estimated from recorded rotor displacements versus static loads applied with the EM actuator on the stationary rotor. The tests begin with the rotor resting on the foil bearings. The upward EM load is then incrementally increased while recording the rotor displacements. Figure 4 illustrates the recorded rotor displacements at the bearing locations versus the applied load. The tests evidence a typical nonlinear relationship denoting first the unloading of the bottom bumps supporting the rotor weight, then the travel of the rotor within the air gap; and for loads larger than ~10 N the loading (stiffening) of the upper bumps in each foil bearing.

Following the method introduced by Ruscitto et al. [13], the estimated radial clearances equal 45 μm and 50 μm (uncertainty ± 2.1 μm) for the bearings at the free and drive ends of the rotor, respectively. Note that this ad-hoc procedure reveals the region where the foil structure is apparently very "soft." It is important to mention that these measurements were conducted after the imbalance response tests were completed. See the section entitled "Posttest Conditions of Foil Bearings" for further details. Unfortunately, the actual clearances for the bearings prior to testing are unknown.

Coastdown Rotordynamics Measurements

Pairs of calibrated imbalance masses (m), ranging from 0.12 g to 0.34 g, are inserted in the test rotor end planes at radius $r = 15.1$ mm. In tests A, the imbalance masses are added at the same angular location at the rotor end planes; whereas in tests B, the imbalance masses are out of phase (180 deg).

The imbalance displacements $u=(2mr/M)=7.4$ μm ($m=0.24$ g) and 10.5 μm ($m=0.34$ g), define the eccentricity of the rotor mass center relative to the center of rotation. Note that a baseline imbalance of ~1 μm is recommended in gas turbine engines operating at a top speed of 25 krpm, G 2.5 balance quality grade [14]. The test imbalances used represent extreme conditions as per API industry standards.

Pressurized air (34.4 kPa gauge) is forced into the end sides of the bearings facing the housing midplane. All tests were conducted with the bearings and rotor at ambient temperature conditions. The feed pressure is small enough to ensure the bearings work under hydrodynamic conditions. When turned on, the motor drives the test rotor to a minimum speed of ~5 krpm. The rig operator increases the motor speed to its top value, 25 krpm, and then proceeds to collect rotor motion data as soon as the motor is turned off and the rotor coasts down to rest. In the tests, the rotor lift-off and touchdown speeds ranged from 2.0 krpm to 5.0 krpm, respectively.

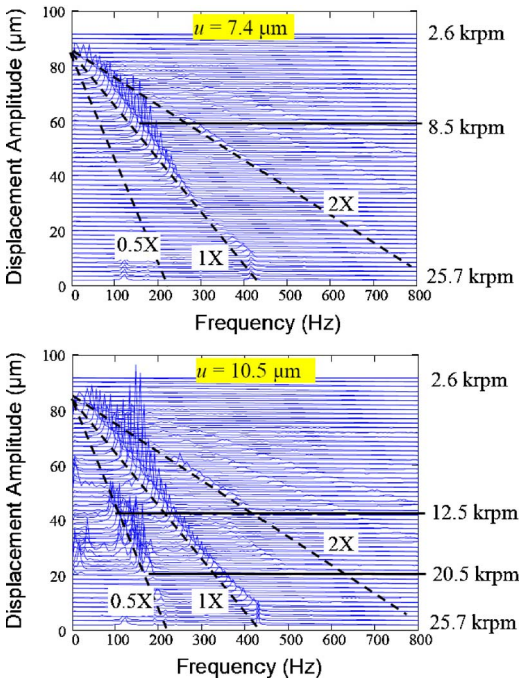


Fig. 5 Waterfalls of rotor coast down response for two in-phase imbalance displacements. Top graph: u=7.4 μm, bottom graph: u=10.5 μm. Air pressure at 34.4 kPa measurements at rotor free end, vertical plane (Y_{FE}).

Figure 5 depicts waterfall plots of the rotor coast down responses for (in phase) imbalance displacements $u=7.4$ μm and 10.5 μm. The horizontal and vertical axes show the frequency content and amplitudes of the rotor motions, while the in-plane axis notes the rotor speed. For the smallest imbalance displacement $u=7.4$ μm, synchronous motion dominates the rotor response during the entire coast down speed range, i.e., 25 krpm to 2 krpm. Figure 6 displays the synchronous and subsynchronous amplitudes of motion and whirl frequency ratio (WFR) versus speed for the test with largest mass imbalance $u=10.5$ μm. The WFR relates the subsynchronous whirl frequency to the shaft spinning frequency.

For the largest imbalance $u=10.5$ μm, subsynchronous motions appear from ~20.5 krpm to 15 krpm with a 50% typical whirl frequency ratio, see Fig. 6. At 17 krpm (~ twice the system critical speed), subsynchronous motions with whirl ratios equal to 0.12, 0.40, and 0.62 denote a transient rub condition with the largest total orbital motion. Below 15 krpm, the subsynchronous whirl motion bifurcates into two whirl ratios, ~0.55 and ~0.45; until disappearing at a shaft speed of ~12.5 krpm. For lower shaft speeds, no subsynchronous motions are observed. At the lowest shaft speed, ~2.6 krpm, multiple whirl ratios, WFRs ~ 0.12, 0.33, and 0.78, denote contact with the foil bearing as the rotor is about to stop. In general, imbalance masses of increasing magnitude have a larger impact on the amplitudes of subsynchronous motion. Also, the rotor speed range with more subsynchronous motion activity occurs at (roughly) two times the system natural frequencies. The recorded rotordynamic performance of the test GFBs shows similar results as in other experimental results [15,16].

For the largest imbalances of each test series, Fig. 7 depicts the recorded subsynchronous amplitudes versus their corresponding

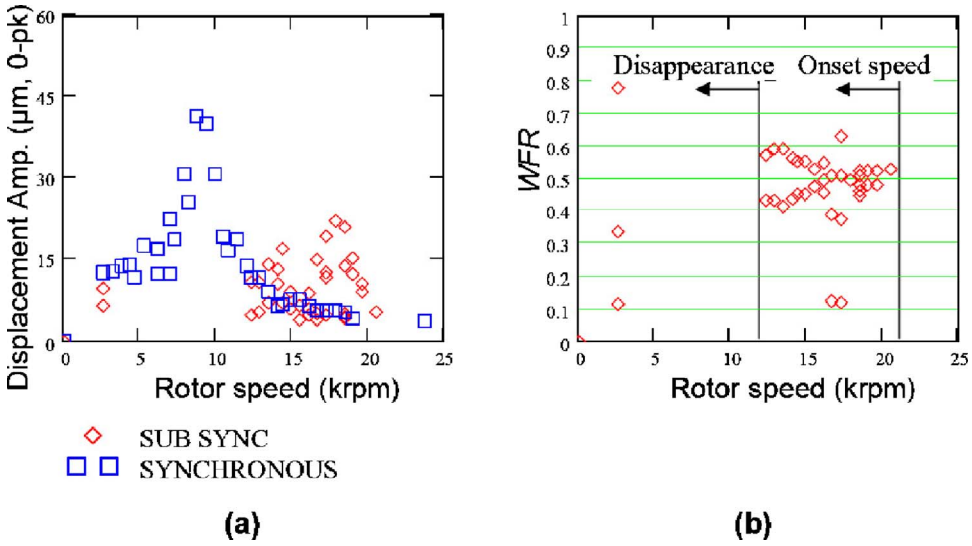


Fig. 6 (*a*) **Amplitudes of synchronous and subsynchronous motion and** (*b*) **whirl frequency ratio (WFR) versus rotor speed for imbalance** u**=10.5** μ**m (in phase, Test A3). Air pressure at 34.4 kPa (5 psig) and measurements at rotor free end, vertical plane (**Y_{FE}**).**

whirl frequencies. The largest amplitudes occur at frequencies equal to $\omega_1 = 105$ Hz (6.3 krpm) and $\omega_2 = 155$ Hz (9.3 krpm). The experimental results demonstrate that the severity of subsynchronous motion is related to the amount of imbalance; i.e., a forced nonlinear phenomenon as opposed to a rotordynamic (self-excited) instability.

Rotordynamic Analysis: Predictions Versus Experimental Results. The test rotor – bearing system is modeled using an in-house linear rotordynamics software. Figure 8 shows the finite element structural model of the test rotor including the connecting shaft and the flexible coupling. The predicted first (free-free) bending mode natural frequency is 3086 Hz, in good agreement with a measured value of 3008 Hz. Note that the connecting shaft is rather flexible. Without this element and the flexible coupling, the first free-free mode natural frequency equals 4096 Hz from rap tests (4100 Hz predicted). Most important, the coupling lateral stiffness equals 1.63×10^5 N/m, of similar magnitude as the GFB structural stiffnesses, affecting not only the rotor/motor alignment but also the rotor dynamic forced response since it is a nonnegligible radial load path.

Foil bearing linearized force coefficients at synchronous frequencies are predicted using the estimated radial clearances and with a structural loss factor (γ) of 0.05 to represent the energy dissipation from dry-friction [17]. Reference [18] details the physical model and computational tool. Note that the bearing (speed) compressibility number, $\Lambda_g = 6\ \mu\Omega(R/C)^2/P_a$ using the nominal clearance equals 0.84 (<1.0) at the top speed, 25 krpm. Thus, the effects of excitation frequency on the GFB force coefficients are not regarded as extremely important on the stability analysis.

An eigenvalue analysis predicts the rotor-GFBs system damped natural frequencies and damping ratios for the test speed range. Figure 9 presents the predicted forward whirl natural frequencies, with critical speeds at 4.25 krpm and 8.75 krpm, and associated to cylindrical and conical modes, respectively. Figure 9 also includes the half frequency whirl line ($0.5X$) with excitation of the natural modes appearing at rotor speeds of 12.25 krpm and 19.25 krpm. The modes are cylindrical and conical with frequencies equal to

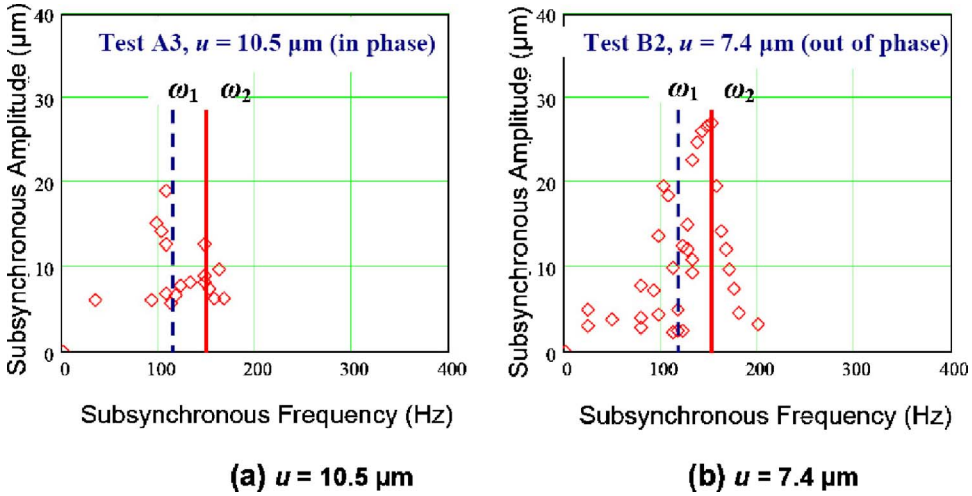


Fig. 7 Subsynchronous amplitudes and respective whirl frequencies for tests with two imbalances, (*a*) u**=10.5** μ**m (in phase) and** (*b*) u**=7.4** μ**m (out of phase)**

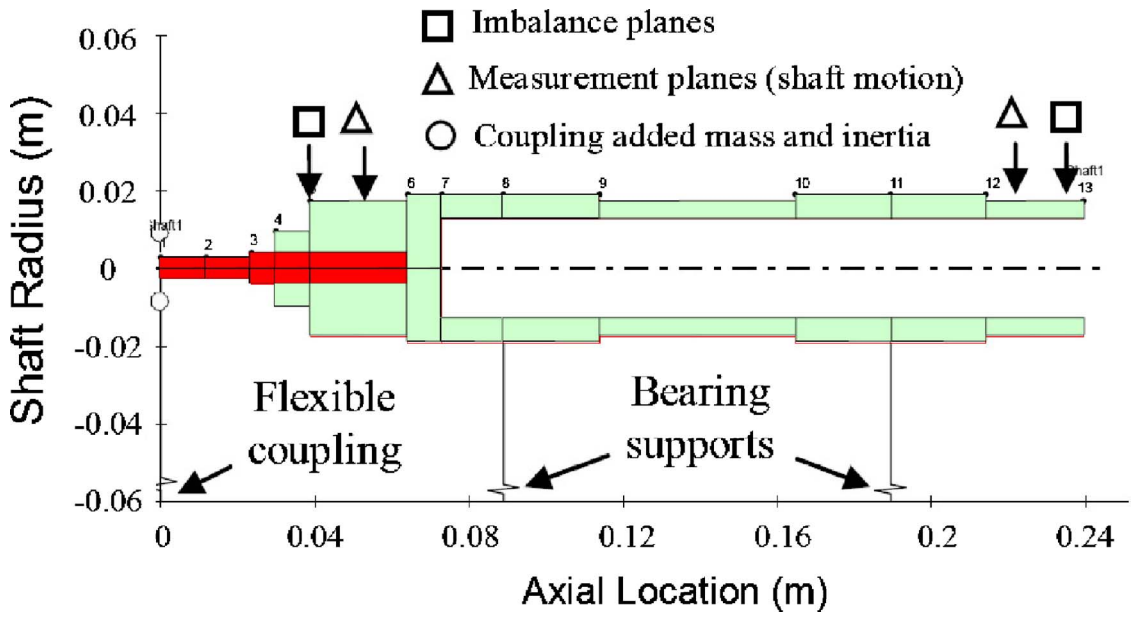


Fig. 8 Finite element model of the test rotor (with connecting shaft and flexible coupling included)

102 Hz (6.12 krpm) and 160 Hz (9.62 krpm), respectively. Note that the test results in Fig. 7 show the subsynchronous frequencies with largest motion amplitudes at $\omega_1 = 105$ Hz (6.3 krpm) and $\omega_2 = 155$ Hz (9.3 krpm), agreeing well with the predictions.

Figure 10 depicts predicted damping ratios decreasing rapidly as rotor speed grows, with the damping for the first mode (cylindrical) becoming negative above 23 krpm, i.e., a rotordynamic instability. The damping ratio for the second (conical) mode is also slightly larger than zero above 23 krpm. Note that the waterfall plots in Fig. 5 show recorded subsynchronous motions with a ~50% whirl frequency for shaft speeds ranging from 20.5 krpm to 15 krpm for the largest imbalance displacement ($u = 10.5$ μm). Thus, the measurements verify the predictions of subsynchronous whirl, albeit at a higher shaft speed.

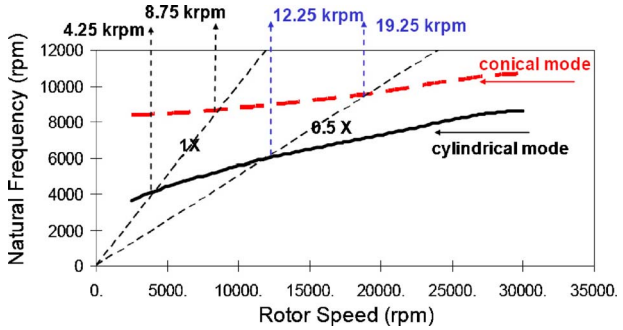


Fig. 9 Predicted damped natural frequencies for rotor–foil bearing system (forward modes)

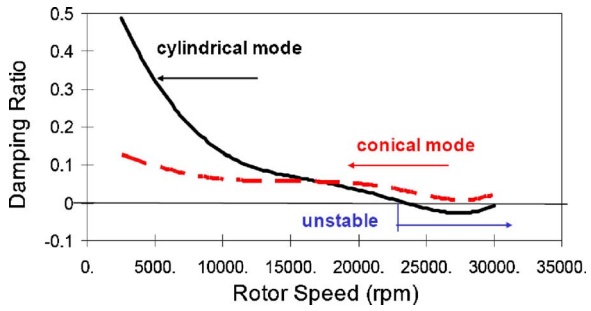


Fig. 10 Predicted damping ratios for rotor–foil bearing system

Figure 11 depicts the measured and predicted amplitudes of synchronous response versus rotor speed for the in-phase imbalance conditions. The top, middle, and bottom graphs correspond to imbalance displacements $u = 7.4$ μm, 9.5 μm, and 10.5 μm, respectively. The recorded rotor coast down responses initiated at 15 krpm since the repeatability of the experimental results was significantly better from this speed rather than from 25 krpm. The lower start speed avoided excessive build up of subsynchronous motions affecting the total rotor motion. For each rotor speed, the baseline condition is subtracted from the measured imbalance response to show only the effect of the imbalance mass used. This procedure is strictly valid in linear systems, however.

In general, the predicted magnitudes agree reasonably well with the test data at the low and high values of rotor speed. Large discrepancies are evident while traversing the system critical speed. From the test results, it is obvious that the test system damping ratio is much lower (~50% or less) than the predicted damping ratios, see Fig. 10. Hence, in the computational GFB model, either the damping coefficients are too large or the cross-coupled stiffnesses too low. The actual reduction in available damping may be characteristic to gas bearings. For example, in Ref. [19] the same conclusion is derived from high speed experiments of a rotor supported on flexure-pivot, tilting pad hybrid gas bearings.

Figure 12 presents *normalized* rotor synchronous response amplitudes versus shaft speed for the three imbalance conditions shown in Fig. 11. The normalization means a multiplication of the response by the smallest imbalance divided by the actual imbalance used. The test data show shifts in the critical speed which decreases from 9 krpm toward 8 krpm as the imbalance mass increases. The test data also evidence an increase in normalized amplitude as the imbalance grows, e.g., characteristic of a nonlinear system. On the other hand, the predictions being based on linearized bearing force coefficients show a unique curve. Thus, the test GFBs have stiffness and damping properties which are amplitude and frequency dependent [17,20,21].

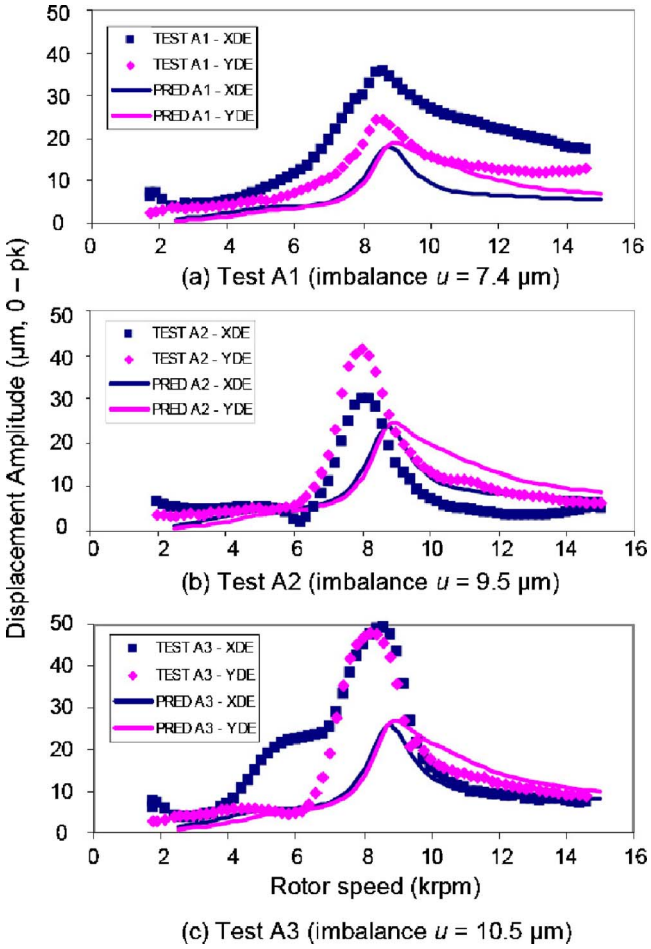


Fig. 11 Measured and predicted amplitudes of synchronous rotor response for (in phase) imbalance tests. (*a*) u=7.4 μm, (*b*) u=9.5 μm, (*c*) u=10.5 μm. Air pressure at 34.4 kPa (5 psig). Measurements at rotor drive end with baseline subtraction, Horizontal (X_{DE}) and vertical (Y_{DE}) planes.

Effect of Air Pressurization on System Response

Tests at constant rotor speed for five feed air pressures ranging from 40 kPa to 340 kPa (6 psig to 50 psig) followed. The forced air is injected at the rig housing midplane and flows into the bearings at one of their axial sides. The test rotor imbalance is u =3.7 μm (in phase condition, m=0.12 g).

In practice, external pressurization is not desirable because of the added expense in power; however, in hot temperature applications it becomes mandatory to maintain the bearings' structural integrity. Note that in the present investigation the gas bearings did not show a measurable temperature rise even for operation with the lowest feed pressures.

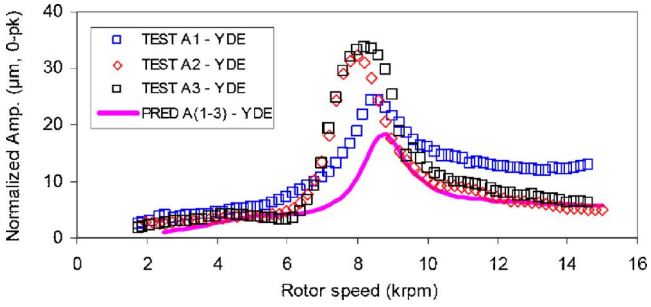


Fig. 12 Normalized amplitudes of rotor synchronous response for imbalance tests (in phase). Air pressure at 34.4 kPa (5 psig). Measurements at rotor drive end, vertical (Y_{DE}) plane, with baseline subtraction.

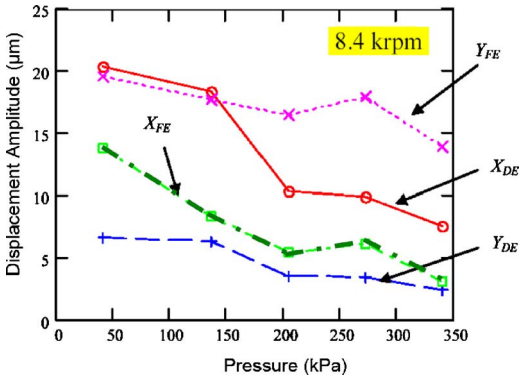


Fig. 13 Amplitudes of synchronous motions at critical speed (8.4 krpm) for increasing air supply pressures and u=3.7 μm

Figure 13 shows, for increasing air supply pressures, the amplitudes of synchronous motion at rotor speeds coinciding with the system critical speed. Air pressurization acts to reduce the rotor synchronous motion amplitude at the critical speeds, thus denoting an increase of damping from the foil bearings. No measurable changes were noted in the system critical speed when increasing the supply pressure, thus eliminating a *Lomakin* annular seal-like effect that would change the bearings' direct stiffness coefficients.

The effect of gas feed pressure on the system stability was evaluated for operation at constant shaft speed (~16 krpm), i.e., about twice the system critical speed. Figure 14 displays FFTs of rotor motion for three increasing supply pressures. There is a notable reduction in the amplitude of subsynchronous motions when increasing the feed pressure. Note that the main subsynchronous frequency, WFR ~ 0.5, splits or bifurcates into two other frequencies as the supply pressure increases.

Rubio [22] argues that the axial flow rate induced by the feed pressure leads to a net reduction of the circumferential flow development along the bearing. This phenomenon decreases cross-coupled stiffnesses and enhances the effective damping from the gas bearings, thus favoring rotor stability. For a clearance of c

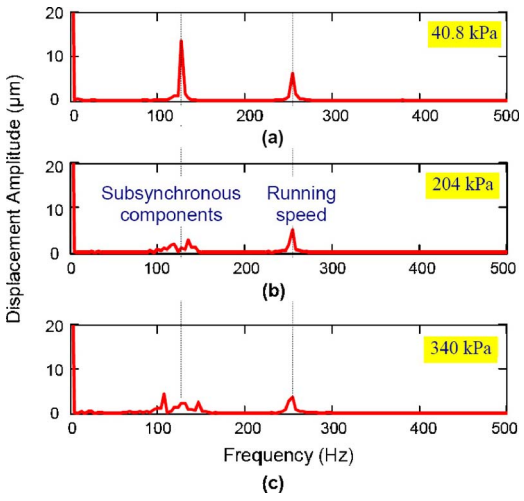


Fig. 14 Amplitude frequency spectrum of rotor motion for three supply pressures; (*a*) 40.8 kPa, (*b*) 204 kPa; and (*c*) 340 kPa. Rotor speed 16 krpm. Measurements at rotor drive end, horizontal plane (X_{DE}). Imbalance, u=3.7 μm.

Fig. 15 Condition of test rotor surface before (top picture) and after (bottom pictures) rotordynamic tests

$=45\ \mu$m (drive end bearing) and shaft speed of 15,000 rpm (one showing severe subsynchronous motions at low feed pressure), the circumferential flow Reynolds number, $\mathrm{Re}_C=(\rho/\mu)\Omega Rc=103$. Conversely, the axial flow Reynolds number $\mathrm{Re}_A=624$ at the highest feed pressure, 340 kPa. Thus, it is likely the circumferential flow development is retarded when $\mathrm{Re}_A > \mathrm{Re}_C$ [22].

Posttest Conditions of Foil Bearings and Rotor. After conducting the exhaustive experimentation [22], the test foil bearings managed to survive severe synchronous and subsynchronous rotor vibrations, by virtue of their inherent flexibility. The top foil presented minor coating wear spots, especially at the bearing edges, while the coating wear in the test rotor was more severe.

Figure 15 shows photographs of the test rotor surface before and after the rotordynamic tests. It is evident that most wear occurred at the location of the journal outboard edges. The coating wear is attributed first to rubbing between journal and top foil at rotor start up and coast down. Predominant conical motions on the test rotor lead to large coating wear at the outboard bearing edges where rotor motions are the largest. The hydrodynamic film pressure dropping to ambient at the bearing edges may also lead to top foil-journal contact at the journal edges.

Figure 16 depicts (negative) photographs of one of the test foil bearings surfaces after the measurements. A white surface represents an unworn top foil coating, while the opaque spots shows wear of the top foil coating. The foil also evidences more wear at the bearing edges, although a few wear spots were found along the direction of applied static load (~135 deg away from the spot weld) and also near the spot weld line.

Conclusions

High performance oil-free turbomachinery implements (GFBs) to improve mechanical efficiency in compact units. Since the GFB design is largely empirical due to their mechanical complexity, the need for experimental demonstration of rotating machinery supported on foil bearing is essential. This work reveals important characteristics of this oil-free bearing technology by conducting rotordynamic experiments of a test rotor supported on two bump-type gas foil bearings.

Tests consisted of recording coast down responses from a top speed of 25 krpm. These tests aid in evaluating the rotordynamic performance of a hollow rotor, 0.98 kg weight, supported on gas foil bearings. Appearance of subsynchronous vibrations is sensitive to the level of imbalance added to the rotor, i.e., the larger the imbalance, the larger the magnitudes of subsynchronous motions. In general, subsynchronous frequencies track the shaft speed and the whirl ratio at the onset of the instability equals 50% of shaft speed. For the largest imbalance condition, amplitudes of subsynchronous vibration are significantly larger than the synchronous ones.

A finite element rotordynamic analysis models the test rotor and connecting shaft; incorporates the coupling lateral stiffness, and uses predicted synchronous bearing force coefficients derived from the static load acting on each gas bearing. Bearing force coefficients rely on the accurate estimation of the actual clearance and the bumps' elastic properties. The rotordynamic analysis predicts a critical speed at 8.5 krpm which correlates well with the experimental result. Calculated damping ratios predict unstable operation at speeds ranging from 17.5 krpm to 25 krpm. The rotor vibration coast down test data show subsynchronous motions with a ~50% whirl frequency ratio for shaft speeds ranging from 20.5 krpm to 15 krpm for the largest test imbalance. In general, the predicted unbalance responses agree well with the test data at the low and high values of rotor speed. Large discrepancies are evident while traversing the system critical speed, except for predictions for the lowest imbalance. A comparison of normalized unbalance amplitudes reveals a nonlinear rotor behavior due to GFB stiffness and damping properties which are amplitude and frequency dependent.

In summary, the flexibility of the coupling and connecting shaft affected the rotor dynamic response, thus making difficult the analysis of the test data and impeding the identification of GFB force coefficients as planned.

External air pressurization through the bearing ends aids to reduce the amplitude of synchronous motions while crossing a critical speed. Feed pressure has no discernible effect on the amplitudes of synchronous motion well above the rotor/bearing system critical speed. Importantly enough, the tests also demonstrate that increasing air pressures ameliorate the amplitudes of subsynchronous motion at shaft speeds roughly two times the critical speed. The increasing axial flow retards the development of the circumferential flow within the gas bearings, thus enhancing the system equivalent damping.

Posttest inspection of the rotor evidenced shaft coating wear at the location of the bearing edges. The top foil presents minor coating wear spots, except at the bearing edges and at the location of static load applications.

The experimental results demonstrate the advantages of GFBs for high speed micropower generation rotating machinery and point out to the need for nonlinear analysis of GFB forced performance. Future experimental work will assess foil bearing performance at higher rotational speeds by adding a faster motor drive system.

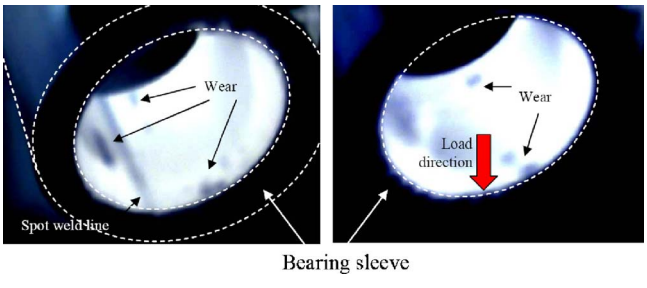


Fig. 16 Condition of foil bearing surface after rotordynamic tests. Wear near the spot weld line (left) and along the load direction (right) in the drive end bearing.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant No. 0322925. The authors also acknowledge the support of the TAMU Turbomachinery Research Consortium (TRC). Thanks to Foster Miller Technologies for providing the test foil bearings.

Nomenclature

c = foil bearing radial clearance (μm)
m = mass imbalance (kg)
M = rotor mass (kg)
r = radius for imbalance mass (m)
u = $2mr/M$, imbalance displacement (μm)
X, Y = horizontal and vertical rotor displacements
ω = frequency (Hz)

Subscripts

DE, FE = drive and free ends

References

[1] Agrawal, G. L., 1997, "Foil Air/Gas Bearing Technology - An Overview," ASME Paper No. 97-GT-347.
[2] Heshmat, H., Walowit, J., and Pinkus, O., 1983, "Analysis of Gas-Lubricated Compliant Journal Bearings," ASME J. Lubr. Technol., **105**(4), pp. 647–655.
[3] Peng, J.-P., and Carpino, M., 1993, "Calculation of Stiffness and Damping Coefficient for Elastically Supported Gas Foil Bearings," ASME J. Tribol., **115**(1), pp. 20–27.
[4] DellaCorte, C., and Valco, M., 2000, "Load Capacity Estimation of Foil Air Bearings for Oil-Free Turbomachinery Applications," STLE Tribol. Trans., **43**(4), pp. 795–801.
[5] Ku, C.-P., and Heshmat, H., 1992, "Complaint Foil Bearing Structural Stiffness Analysis—Part I: Theoretical Model - Including Strip and Variable Bump Foil Geometry," ASME J. Tribol., **114**(2), pp. 394–400.
[6] Ku, C.-P., and Heshmat, H., 1993, "Complaint Foil Bearing Structural Stiffness Analysis—Part II: Experimental Investigation," ASME J. Tribol., **113**(3), pp. 364–369.
[7] Iordanoff, I., 1999, "Analysis of an Aerodynamic Complaint Foil Thrust Bearing: Method for a Rapid Design," ASME J. Tribol., **121**, pp. 816–822.
[8] Ku, C.-P., and Heshmat, H., 1994, "Structural Stiffness and Coulomb Damping in Compliant Foil Journal Bearing: Theoretical Considerations," STLE Tribol. Trans., **37**(3), pp. 525–533.
[9] Ku, C.-P., and Heshmat, H., 1994, "Structural Stiffness and Coulomb Damping in Compliant Foil Journal Bearing: Parametric Studies," STLE Tribol. Trans., **37**(3), pp. 455–462.
[10] DellaCorte, C., and Valco, M., 2003, "Oil-Free Turbomachinery Technology for Regional Jet, Rotorcraft and Supersonic Business Jet Propulsion Engines," AIAA Paper No. ASABE–2003–1182.
[11] Gu, A., 1988, "Process Fluid Foil Bearing Liquid Hydrogen Turbopump," AIAA Paper No. 88-3130.
[12] Chen, H., Howarth, R., Geren, B., Theilacker, J., and Soyars, W., 2001, "Application of Foil Bearings to Helium Turbocompressor," *Proceedings of the 30th Turbomachinery Symposium*, Houston, TX, pp. 103–112.
[13] Ruscitto, D., Mc Cormick, J., and Gray, S., 1978, "Hydrodynamic Air Lubricated Compliant Surface Bearing For An Automotive Gas Turbine Engine I-Journal Bearing Performance," NASA CR-135368.
[14] Rao, J., S., 1983, *Rotor Dynamics*, New Age International (P) Limited, New Delhi, India, pp. 364–369.
[15] Swanson, E., Walton, J. F., and Heshmat, H., 2002, "A Test Stand for Dynamic Characterization of Oil-Free Bearings for Modern Gas Turbine Engines," ASME Paper No. GT2002-30005.
[16] Lee, Y. B., Kim, T. H., Kim, C. H., and Lee, N. S., 2003, "Suppression of Subsynchronous Vibrations Due to Aerodynamic Response to Surge in a Two-Stage Centrifugal Compressor With Air Foil Bearings," STLE Tribol. Trans., **46**, pp. 428–434.
[17] Rubio, D., and San Andrés, L., 2007, "Structural Stiffness, Dry-Friction Coefficient and Equivalent Viscous Damping in a Bump-Type Foil Gas Bearing," ASME J. Eng. Gas Turbines Power, **129**(1), pp. 1–9.
[18] Kim, T. H., and San Andrés, L., 2005, "Heavily Loaded Gas Foil Bearings: A Model Anchored to Test Data," ASME Paper No. GT2005-68486.
[19] Zhu, S., and San Andrés, L., 2004, "Rotordynamic Performance of Flexure Pivot Hydrostatic Gas Bearings for Oil-Free Turbomachinery," ASME Paper No. GT2004-53621.
[20] Heshmat, H., and Ku, C.-P., 1994, "Structural Damping of Self-Acting Compliant Foil Journal Bearings," ASME J. Tribol., **116**, pp. 76–82.
[21] Lee, Y. B., Kim, T. H., Kim, C. H., Lee, N. S., and Choi, D. H., 2004, "Dynamic Characteristics of a Flexible Rotor System Supported by a Viscoelastic Foil Bearing (VEFB)," Tribol. Int., **37**, pp. 679–687.
[22] Rubio, D., 2005, "Rotordynamic Performance of a Rotor Supported on Bump-Type Foil Bearings: Experiments and Predictions," M.S. thesis, Texas A&M Univ., College Station, TX.

Identification of Force Coefficients in a Squeeze Film Damper With a Mechanical End Seal—Part I: Unidirectional Load Tests

Luis San Andrés
e-mail: lsanandres@mengr.tamu.edu

Adolfo Delgado
e-mail: adelgam@tamu.edu

Mechanical Engineering Department,
Texas A&M University,
College Station, TX 77843

Squeeze film dampers (SFDs) with low levels of external pressurization and poor end sealing are prone to air entrapment, thus not generating enough damping capability. Single frequency, unidirectional load tests were conducted on a SFD test rig replicating a commercial jet-engine configuration. The damper journal is 2.54 cm *in length and* 12.7 cm *in diameter, with nominal clearance of* 0.127 mm*. The SFD feed end is flooded with oil, while the discharge end contains a recirculation groove and four orifice ports, and a mechanical seal ring in contact with the damper journal. A wave spring pushes the ring ensuring tight sealing to prevent gas ingestion. The mechanical seal also serves to contain the lubricant within the squeeze film land for extended periods of time and; while in operation, to prevent contamination of the ball bearing cartridge. The measurements conducted without and with lubricant in the squeeze film lands, along with a frequency domain identification procedure, render the mechanical seal dry-friction force and viscous damping force coefficients as functions of frequency and motion amplitude. The end seal arrangement is quite effective in eliminating side leakage and preventing air entrainment into the film lands. Importantly enough, the dry friction force, arising from the contact forces in relative motion, increases significantly the test element equivalent viscous damping coefficients. The identified system damping coefficients are thus frequency and motion amplitude dependent, albeit decreasing rapidly as the motion parameters increase. Identified squeeze film force coefficients, damping and added mass, agree well with predictions based on the full film, short length damper model.*
[DOI: 10.1115/1.2436571]

Introduction

Squeeze film dampers (SFDs) are widely used to provide external damping in aircraft gas turbines supported on rolling element bearings, as well as in hydrocarbon compressors, for example. SFDs also aid to attenuate rotor synchronous responses at passage through critical speeds where rotor vibration amplitude due to imbalance peaks. The forced performance (damping capability) of a SFD depends on its geometrical configuration and lubricant viscosity, as well as on its operational parameters such as the flow regime, kinematics of journal motion, lubricant cavitation type, and air ingestion and entrapment among others [1].

Della Pietra and Adilleta [2,3] present a monumental review of the experimental and analytical research conducted in SFDs over the last 40 years. The reviews include descriptions of oil lubricated dampers and end seal types, the advances in modeling and parameter identification procedures, and a discussion on the moderate capability of predictive models when compared to the extensive body of experimental results. San Andrés [4] discusses issues of importance in the design and operation of SFDs, outlining distinctions between laboratory experimental results mostly related to fully flooded SFD configurations and actual practice which demands low levels of external pressurization and end conditions open to the environment.

Presently, with increasing rotor flexibilities and shaft speeds, high-performance turbomachinery undergoes high dynamic loads and large displacements. Under these conditions, air ingestion and entrapment deteriorates the forced performance of open ends and partially sealed SFDs incorporating end plate seals, piston rings or O-rings. Fully sealed SFDs are less prone to the effects of air ingestion; however this mechanically complex configuration is difficult to attain in practice [4].

Levesley and Holmes [5] compare the damping capacity and general performance of various SFD configurations, including various types of end seals. The tests results show that piston-ring end sealed SFDs have a better performance over two other end seal arrangements tested (end-chambered and end plate seals). In practice, however, piston rings are not favored due to their tendency to lock the damper when not properly designed and installed.

In a rotordynamics test rig, De Santiago and San Andrés [6] estimate the damping coefficients of an integral SFD with end plate seals from measured rotor responses to calibrated imbalances as well as from impact loads exerted on the rotor. In the experiments, the clearance or gap of each end plate seal was tightly controlled. The test measurements and identification show that the damping coefficients of the end sealed damper (0.0762 mm clearance) are twice as large as those from the same damper with open ends. Most importantly, while traversing two critical speeds, the recorded rotor motions are proportional to the

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received July 18, 2006; final manuscript received July 25, 2006. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2006: Land, Sea and Air (GT2006), Barcelona, Spain, May 8–11, 2006. Paper No. GT2006-90782.

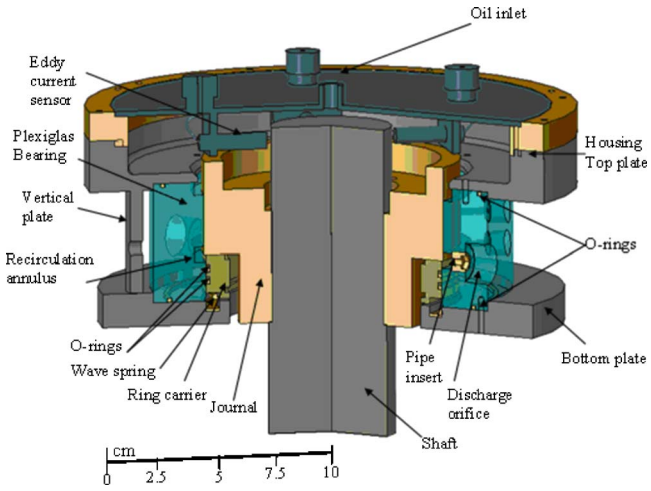


Fig. 1 Assembly cut view of SFD with mechanical seal

imbalance, independently of its magnitude. Hence, the integral SFDs evidenced a linear behavior, thus opposing the generally accepted view that SFDs are highly nonlinear mechanical elements.

Arauz and San Andrés [7] conducted measurements of dynamic film pressures in a constrained orbit centrally grooved SFD with a serrated piston ring at the damper discharge end. This unique type of piston ring makes uniform the end oil leakage while adding a mechanical resistance to prevent air ingestion. The tests evaluated three increasing groove depths for journal circular motions of radius ~50% of the film clearance. For small to moderate groove depths, the measurements evidence dynamic pressures at the groove similar in magnitude as those recorded in the film lands. Consequently, as originally advanced by San Andrés and Vance [8], relatively large damping forces are generated at the central groove, which effectively contribute to the overall damping ability of the SFD. Recently, Kim and Lee [9] present further experiments to assess the effects of a central groove and end seals on the damping capability of a SFD. Comparison of experimentally derived damping coefficients to predictions from a model based on the short length bearing theory modified with an end seal condition are reasonable. However, predicted added mass coefficients are a fraction of the experimentally derived coefficients.

Since 1985, research at the authors' laboratory has steadily advanced the technology of SFDs and their industrial application. Recent efforts elucidated the issue of gas entrainment and entrapment; advancing simple engineering models for reliable prediction of open-ends SFD forced performance, see Refs. [10,11] for example. The research also focuses on the development of accurate parameter identification techniques with analysis of the test data in the frequency domain [12]. In spite of the progress advanced, there are still issues yet to be fully understood. One issue is the poor correlation of predicted and experimentally derived added mass coefficients; theory typically under predicting by ~50% the measurements.[1] Another issue relates to end sealing, with some actual SFD configurations showing extreme complexity and little experimental results to verify their predicted performance.

Presently, the test rig in Ref. [10] accommodates a mechanical end seal ring that pushes into the damper journal thus avoiding in practice the need for a centering (external) spring. The mechanical seal introduces a dry-friction force that adds to the squeeze film forces, thus complicating the identification of the SFD force coefficients. Figure 1 depicts a schematic cutaway of the test SFD closely replicating an industrial configuration. The damper housing includes a deep circular groove at the damper outlet closest to the mechanical seal. The damper can operate as an open end, partially sealed, or fully sealed, end configuration, depending on the normal force upon assembly of the mechanical seal. The SFD designed feature is full containment of the lubricant within the fluid film for long periods of storage. No prior experimental work on this type of damper is available in the open literature.

Experiments aiming to identify the overall damping capability of the mechanically sealed SFD follow. The parameter identification relies on frequency domain methods to extract stiffness and equivalent viscous damping coefficients from measured mechanical impedances (force/displacement).

Test Rig Description

Figure 2 depicts the test rig comprising a vertical rigid shaft mounted on three precision ball bearings (natural frequency 400 Hz [10]) and holding a steel journal 127 mm (5 in.) in diameter and 76.2 mm (3 in.) in length. The bearing assembly (see Fig. 1), consists of an acrylic ring sandwiched between two steel plates. Thick steel beams connect the ring and plates and also serve as an interface to apply external forces onto the bearing assembly. Four steel rods connect the composite bearing housing to a top structure. The rods provide a structural stiffness to the test bearing section. On the top structure, a mechanism comprising two flat plates, sliding relative to each other, allows centering and off-centering positioning of the bearing with respect to its journal.

The bearing housing design integrates a SFD land and a discharge recirculation annulus. The SFD radial clearance (c) is 127 μm$\pm$6 μm at the test temperature (23°C). The axial film land length (L) is 25.4 mm and the circular end groove is 4.25 mm in depth and 7.62 mm in length.

A steel ring, located at the bottom of the bearing assembly, along with the journal bottom surface provides a metal–metal mechanical seal. A wave spring, pushing the ring holder against the journal, applies a contact force between the matting surfaces. The oil is fed from the top and exits the damper at the recirculation annulus through four outlet ports. The discharge flow area is regulated by flow restrictors, each with a hole of 2.8 mm in diameter.

The mechanical seal must contain the lubricant within the film for long periods of storage. While in operation the mechanical seal should prevent oil leakage that may contaminate the ball bearing cartridge for the application intended. Note that the mechanical seal, because of its assembly, introduces a frictional force which should neither be too large to lock up the damper preventing its operation nor too low to permit oil leakage.

Two electromagnetic shakers (maximum 448 N), suspended from separate steel frames, deliver excitation forces onto the test device (x and y directions). Slender steel stingers connect the shakers to the bearing housing. The test rig instrumentation includes two piezoelectric load cells fastened to the bearing and the end of a stinger, two accelerometers attached to the bearing, and four eddy current sensors attached to the bearing assembly and facing the journal end. Other ancillary instrumentation includes pressure sensors at the lubricant inlet section, thin-film land and recirculation annulus, as well as flow meters at the damper inlet plenum and return line into the oil sump. Several thermocouples record the lubricant temperature at the inlet, film, and discharge ports.

The experiments were conducted with an ISO VG 2 lubricant of nominal viscosity equal to 1.68 cP and 0.72 cP at temperatures 40°C and 100°C, respectively. The lubrication system includes a delivery pump and oil heating element, supply and return lines, and static pressure gauges at the inlet section and recirculation groove (discharge plane).

Methodology for System Parameter Identification

The test system is complex and nonlinear since the mechanical seal introduces a dry-friction force which depends not only upon

[1]This deficiency does not just occur in SGDs but also in liquid annular seals and hydrostatic bearings as the extensive literature attests.

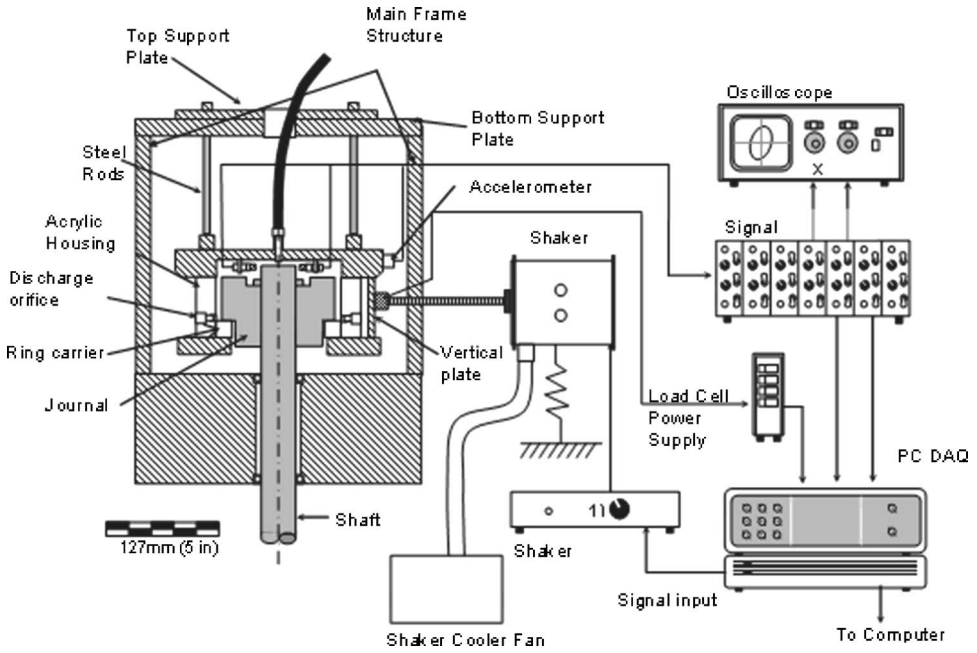


Fig. 2 Test rig for dynamic force measurements and flow visualization in a sealed end SFD

the assembly normal force and friction coefficient of the mating surfaces but also on the amplitude and frequency of the dynamic motions of the damper bearing. The aim is to assess the squeeze film damper forces uniquely. A proper methodology must be in place to achieve this end. First, impact tests delivered on the test rig without the mechanical seal in place and no lubricant (dry condition) in the damper serve to identify the bearing mass, structural stiffness, and residual or remnant (viscous) damping coefficients. Next, the mechanical seal is assembled, and under dry conditions (no lubricant), single frequency loads are exerted on the system, and motions recorded. The identification procedure relies on equating the total external work to the energy dissipated by dry-friction forces. The estimated dry-friction force can be equated to an equivalent viscous damping action using a well known engineering formula.

Finally, the test section is filled with lubricant and the feed pump supplies oil at controlled conditions in pressure and temperature. The effectiveness of the mechanical seal is assessed through in and out flow measurements. Next, single frequency dynamic loads are exerted on the lubricated system and bearing motions recorded. Load/velocity transfer functions in the frequency domain are constructed for estimation of the system viscous damping and dynamic stiffness coefficients. A final post-test processing subtracts the damping contribution of the mechanical seal to reveal the viscous damping coefficients from the squeeze film damper alone. It is worth noting that in the experiments, the shaker loads are controlled for the test system to show characteristic motions, e.g., circular centered orbits of fixed amplitude, for example.

Test System Stiffness, Mass, and Remnant Viscous Damping Coefficients

The first set of experiments consists of static and impact tests conducted on the unlubricated (dry) test device without the end seal in place. The equations of motion for this system configuration are

Table 1 SFD test section identified parameters (no lubricant)

Parameter		x	y
Stiffness	K_s (kN/m)	788 (±40)	823 (±46)
Mass	M (kg)	9.7 (±0.5)	
Damping	C_{rv} (N s/m)	367	315
Damping ratio	ζ	0.07	0.05
Natural frequency	f_n (Hz)	47±1	49±1
(goodness of fit)	R^2	0.96	0.98

$$M\begin{Bmatrix}\ddot{x}\\ \ddot{y}\end{Bmatrix}+\begin{Bmatrix}C_{rv_x}\dot{x}\\ C_{rv_y}\dot{y}\end{Bmatrix}+\begin{Bmatrix}K_{s_x}x\\ K_{s_y}y\end{Bmatrix}=\begin{Bmatrix}F_{x(t)}\\ F_{y(t)}\end{Bmatrix} \quad (1)$$

where M is the bearing effective mass; K_s is the structural (rods) stiffness; and C_{rv} is a residual or remnant "viscous" damping coefficient. Table 1 presents the identified parameters, along with their associated uncertainty. Delgado [13] details the experiments, identification procedure, and uncertainty analysis.

The stiffness along the two directions are not exactly identical due to a slight orthotropy in the test rig. The residual damping ratio (ζ) is typical of steel structures.

Estimation of Dry Friction Force and Equivalent Viscous Damping Coefficient

After assembling the mechanical end seal, single frequency dynamic loads are applied to induce circular centered orbits. The aim is to asses the effect of the mechanical seal on the system damping. The tests comprised of load magnitudes[2] of ~34 N and ~40 N with frequencies ranging from 20 Hz to 200 Hz. The excitation loads exerted along directions x and y are of the same magnitude but 90 deg out of phase (i.e., circular load). The equations of motion for the dry-test bearing with mechanical seal are

[2]Loads of lower magnitude could not induce motion in the test system. Sometimes for the lowest load, stick-slip occurred over certain frequency ranges, see Ref. [13] for details.

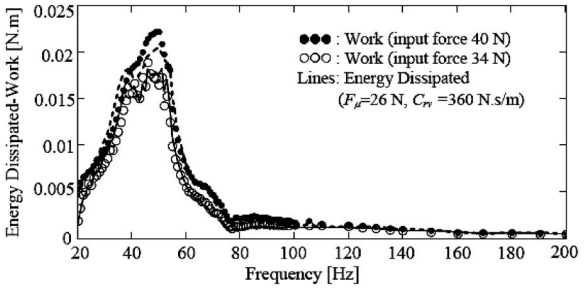


Fig. 3 Work exerted by input force (=dissipated energy) estimated from combined damping model (dry SFD, end seal in place)

$$M\begin{Bmatrix}\ddot{x}\\ \ddot{y}\end{Bmatrix}+\begin{Bmatrix}C_{\mathrm{rv}_x}\dot{x}\\ C_{\mathrm{rv}_y}\dot{y}\end{Bmatrix}+\begin{Bmatrix}K_{s_x}x\\ K_{s_y}y\end{Bmatrix}=\begin{Bmatrix}F_{x(t)}\\ F_{y(t)}\end{Bmatrix}-\begin{Bmatrix}F_x\\ F_y\end{Bmatrix}_{\mathrm{seal}} \tag{2}$$

The mechanical seal reaction force is expressed as

$$\begin{Bmatrix}F_x\\ F_y\end{Bmatrix}_{\mathrm{seal}}=F_\mu\begin{Bmatrix}\dfrac{\dot{x}}{|\nu|}\\ \dfrac{\dot{y}}{|\nu|}\end{Bmatrix}\approx\begin{Bmatrix}C_{\mathrm{seal}_x}\dot{x}\\ C_{\mathrm{seal}_y}\dot{y}\end{Bmatrix} \tag{3}$$

where F_μ is the dry-friction force and $\nu=(\dot{x},\dot{y})$ is the velocity vector. C_{seal} is an equivalent "viscous" damping coefficient that follows from an equivalence of the energy dissipated by viscous forces to the energy from dry friction forces over one period of motion ($T=2\pi/\omega$) [14]

$$C_{\mathrm{seal}}\int\nu^2dt=E_{\mathrm{DRY}}=F_\mu\int\nu\,\mathrm{sign}(\nu)dt\rightarrow C_{\mathrm{seal}}=\frac{F_\mu}{\omega|x_\alpha|};\quad\alpha=x,y \tag{4}$$

For periodic motions $x=\bar{x}e^{i\omega t}=|x|e^{i(\phi_1+\omega t)}$; $y=\bar{y}e^{i\omega t}=|y|e^{e^{i(\phi_2+\omega t)}}$, the work ($W$) exerted by the external force over one period of motion is

$$W=\pi[|F_x||x|\sin(\phi_1)+|F_y||y|\sin(\phi_2)] \tag{5}$$

and must equal the energy dissipated by the nonconservative elements in the system, mainly dry friction, i.e., E_{dry}. Above, $(|F_x|,|F_y|)$ and $(|x|,|y|)$ are the magnitudes of the applied forces and recorded motions; and ϕ_1 and ϕ_2 are the phase lags of the motion response relative to the forces acting along the x and y directions, respectively.

The energy analysis of the test results evidences a dry friction force (F_μ) equal to 26 N and a remnant viscous damping coefficient (C_{rv})=360 N s/m, nearly identical to the ones obtained from the tests without the mechanical seal in place, (see Table 1). The friction force relates to the magnitude of the contact force at the mechanical seal interface, estimated at 70 N (±10 N) from a direct measurement using a dynamometer. This force is the minimum necessary to seal the damper, i.e., avoid side oil leakage, as determined from several trial assembly tests. The friction coefficient (μ) of the seal mating surfaces is ~0.37, typical of steel–steel surfaces in contact [15].

Figure 3 shows the external work from the applied forces and the mechanical energy dissipation evaluated with the estimated dry friction force and remnant viscous damping for the nonlubricated system. The model correlation is very good, even though the system shows stick–slip at certain frequencies. Figure 4 depicts the equivalent viscous damping coefficient combining the contributions of the mechanical seal (C_{seal}) and remnant (structural) damping. The residual damping coefficient (C_{rv}) is but a small fraction compared to the equivalent "viscous" action of the dry friction effect. Note that the "viscous" damping (C_{seal}) arising from dry friction at the contact surface is lowest at frequencies from 40 Hz to 50 Hz, which are precisely those at which the system peaked in motion; i.e., natural frequencies. More importantly, however is the verification that the "viscous" damping from the mechanical seal is nonlinear, i.e., frequency and amplitude dependent. Obviously, the "damping" effect of the mechanical seal is more pronounced at low frequencies.

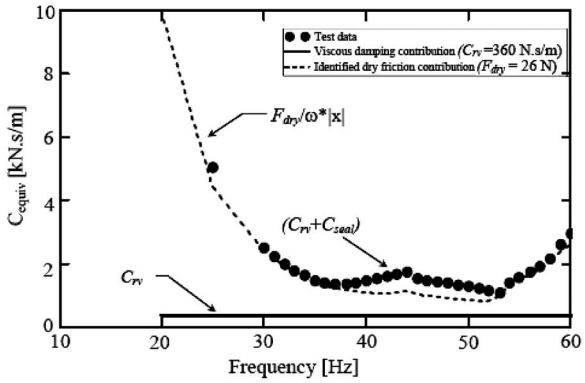


Fig. 4 Equivalent viscous damping (dry friction+residual) versus excitation frequency (test-Dry SFD, end seal in place)

Estimation of Viscous Damping Coefficients for Lubricated System and Identification of Squeeze Film Damping and Inertia Force Coefficients

Once the end seal and structural parameters are characterized, lubricant flows through the squeeze film lands, and *unidirectional* excitation loads of single frequency ranging from 20 Hz to 60 Hz, along the x and y orthogonal directions follow. The tests include five increasing motion amplitudes (13–39 μm) at each excitation frequency. The bearing is initially centered within its journal. The data acquisition (DAQ) system is set to automatically adjust the excitation load amplitude at each frequency to maintain the preset bearing displacements throughout the test frequency range. Maintaining a constant dynamic displacement amplitude favors the identification of parameters in a mechanical systems with dry friction [16].

The lubricant flowing through the damper is under laminar flow conditions with inlet and discharge pressures set to 1.16 bar and 1.06 bar, respectively, and temperature equal to 24 °C. The measured flow rate is proportional to the pressure drop across the squeeze film land and clearance (c^3), and inversely proportional to the oil viscosity (η) and film length (L). The leakage through the contact surfaces is insignificant, thus evidencing the effectiveness of the mechanical seal. Delgado [13] details the damper through flow measurements for a variety of feed pressures and lubricant at increasing temperatures.

The equations of motion for the test system including the lubricated SFD and mechanical seal are

$$(M+M_f)\begin{Bmatrix}\ddot{x}\\ \ddot{y}\end{Bmatrix}+\begin{Bmatrix}C_{\mathrm{rv}_x}\dot{x}\\ C_{\mathrm{rv}_y}\dot{y}\end{Bmatrix}+\begin{Bmatrix}K_{s_x}x\\ K_{s_y}y\end{Bmatrix}=\begin{Bmatrix}F_{x(t)}\\ F_{y(t)}\end{Bmatrix}-\begin{Bmatrix}F_x\\ F_y\end{Bmatrix}_{\mathrm{seal}}-\begin{Bmatrix}F_x\\ F_y\end{Bmatrix}_{\mathrm{SFD}} \tag{6}$$

where M_f (0.62 kg±0.01 kg) represents the mass of fluid enclosed in the plenum above the squeeze film land and in the recirculation annulus. The SFD reaction forces follow the linearized description

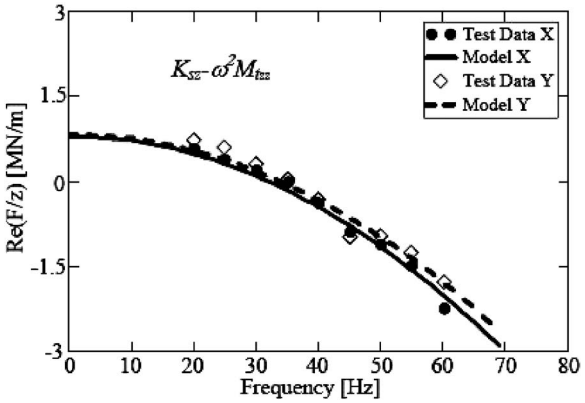


Fig. 5 Dynamic stiffnesses from unidirectional load (single frequency) excitation tests and analytical model (amplitude of motion 39 μm, K_{sx}=788 kN/m, K_{sy}=823 kN/m, M_{txx}=19.7, M_{tyy} =18.4, lubricated SFD)

$$\begin{Bmatrix} F_x \\ F_y \end{Bmatrix}_{\text{SFD}} = \begin{bmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{bmatrix}_{\text{SFD}} \begin{Bmatrix} \dot{x} \\ \dot{y} \end{Bmatrix} + \begin{bmatrix} M_{xx} & M_{xy} \\ M_{yx} & M_{yy} \end{bmatrix}_{\text{SFD}} \begin{Bmatrix} \ddot{x} \\ \ddot{y} \end{Bmatrix} \quad (7)$$

where $\{C_{\alpha\beta}\}_{\alpha\beta=x,y}$, $\{M_{\alpha\beta}\}_{\alpha\beta=x,y}$ are the squeeze film damping and inertia force coefficients, respectively. Recall that a SFD does not generate stiffness coefficients. Furthermore, as determined from the experimental results, the cross-coupled damping and added mass terms are negligible since the end seal effectively prevents air entrapment into the damper at the tested frequencies. Thus, the x and y motions are uncoupled. For prediction purposes, the 2π film (uncavitated) SFD model should be appropriate. As before, the effects of dry friction are modeled with an equivalent viscous damping coefficient (C_{seal}), which for unidirectional motions equals [14]

$$C_{\text{seal}_\alpha} = \frac{4F_\mu}{\pi\omega|x_\alpha|}; \quad \alpha = x,y \quad (8)$$

where x_α denotes displacements along the x or y directions. With this consideration, the equations of motion become linear. The applied dynamic forces and ensuing dynamic motions are of unique frequency (ω) and expressed as

$$F_\alpha(t) = F_{\alpha c}\cos(\omega t) + F_{\alpha s}\sin(\omega t) = (F_{\alpha c} - iF_{\alpha s})e^{i\omega t} = \bar{F}_\alpha e^{i\omega t}; \quad \alpha = x,y \quad (9a)$$

$$x_{(t)} = (x_c - ix_s)e^{i\omega t} = \bar{x}e^{i\omega t} = |x|e^{i(\phi_1+\omega t)}$$

$$\dot{x} = i\omega\bar{x}e^{i\omega t}; \quad \ddot{x} = -\omega^2\bar{x}e^{i\omega t} \quad (9b)$$

similarly for $y(t)$. Thus, in the frequency domain the equations of motion turn into the algebraic expressions

$$[(K_{s_\alpha} - \omega^2 M_{t_{\alpha\alpha}}) + i\omega C_{t_{\alpha\alpha}}]\bar{x}_\alpha = \bar{F}_\alpha; \quad \alpha = x,y \quad (10)$$

where

$$M_{t_{\alpha\alpha}} = M + M_f + M_{\text{SFD}_{\alpha\alpha}}$$

and

$$C_{t_{\alpha\alpha}} = C_{\text{SFD}_{\alpha\alpha}} + C_{\text{seal}_\alpha} + C_{\text{rv}_\alpha}, \quad \alpha = x,y \quad (11)$$

represent the system total mass and the equivalent viscous coefficients, a superposition of squeeze film plus dry friction plus remnant structural damping. The system damping and dynamic stiffness parameters can be readily identified from Eq. (10) as

Table 2 Identified system mass and squeeze film inertia force coefficients. Unidirectional single frequency load tests (amplitude of motion: 38 μm, frequency range 20–60 Hz).

Parameter		*xx*	*yy*
Total identified mass	(M_t) (kg)	19.7	18.4
SFD inertia coefficient	(M_{SFD}) (kg)	**9.4**	**8.1**
R^2		0.98	0.96
Predicted added mass	(M_{SFD}) (kg)	8.2	

$$C_{t_{\alpha\alpha}} = \frac{1}{\omega}\,\text{Im}\left(\frac{\bar{F}_\alpha}{\bar{x}_\alpha}\right); \quad (K_{s_\alpha} - \omega^2 M_{t_{\alpha\alpha}}) = \text{Re}\left(\frac{\bar{F}_\alpha}{\bar{x}_\alpha}\right); \quad \alpha = x,y \quad (12)$$

Figure 5 shows the test derived dynamic stiffness of the lubricated system, $(K - M\omega^2)$, for the 38 μm motion amplitude tests, including predictions using the identified structural stiffness, (see Table 1). Table 2 presents the estimated total system mass coefficients and the derived SFD inertia coefficients ($M_{\text{SFD}_{xx}}, M_{\text{SFD}_{yy}}$) after subtraction of $(M+M_f)$=10.3 kg, i.e., the bearing housing and enclosed fluid mass in the plenum and recirculation groove. The table includes a prediction of the added inertia for an uncavitated squeeze film of length ($2L$), i.e.

$$M_{\text{SFD}_{xx}} = M_{\text{SFD}_{yy}} = \frac{1}{2}\frac{\rho\pi R(2l)^3}{10c} \quad (13)$$

which compares well with the test values. The typical formula given in the archival literature [4] predicts a too low value when compared to the test data, see also Ref. [10]. The type of inlet condition in the test damper, i.e., no allowance for back flow, determines selecting a length twice as long as the actual one [17].

Figure 6 depicts the system damping coefficients ($C_{t_{xx}}, C_{t_{yy}}$) identified at each excitation frequency and for two amplitudes of dynamic displacement, 13 μm and 38 μm, top, and bottom graphs, respectively. Recall that the system damping combines the

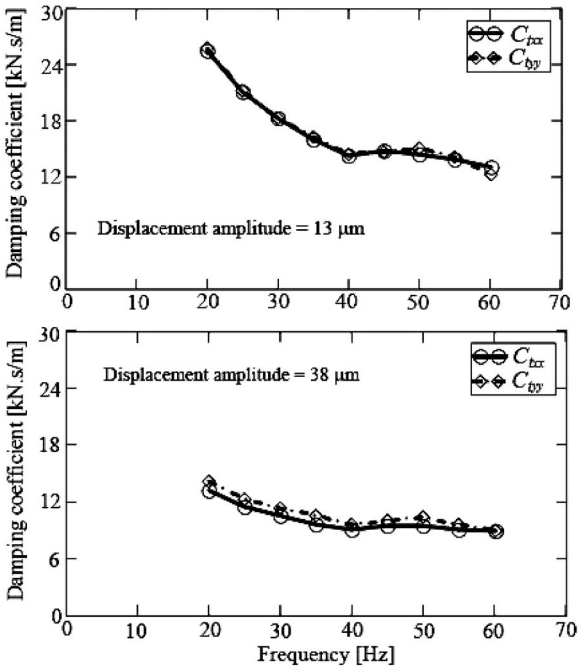


Fig. 6 Identified system damping coefficients ($C_{t_{xx}}, C_{t_{yy}}$) versus excitation frequency. Unidirectional load tests (lubricated SFD).

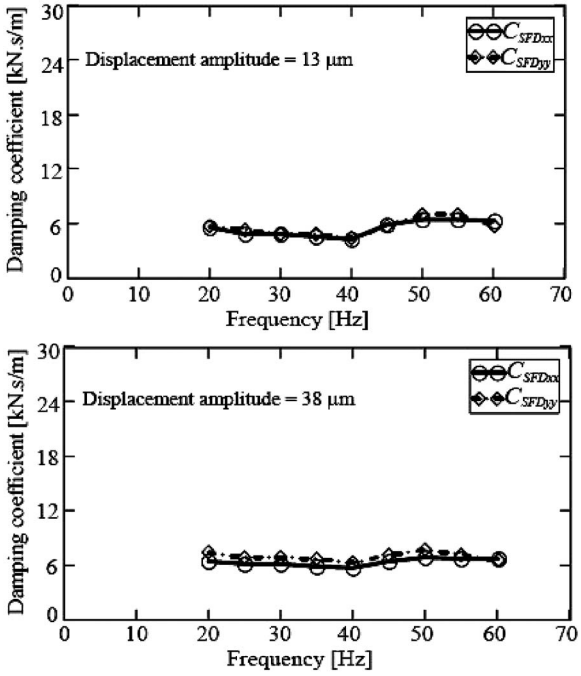


Fig. 7 Derived squeeze film damping coefficients ($C_{SFD_{xx}}, C_{SFD_{yy}}$) versus excitation frequency. Unidirectional load tests (lubricated SFD).

dry friction from the mechanical seal, the viscous squeeze film damping, and the remnant damping action from the structure. The experimental results show that the system is highly over damped. The "viscous" damping arising from dry friction is inversely proportional to the amplitude of motion and excitation frequency, i.e., $C_{\text{seal}} \sim F_\mu/(\omega|x|)$. Therefore, viscous damping from the mechanical seal is dominant at low frequencies and small amplitude of motions. On the other hand, the squeeze film damping coefficient is expected to be independent of excitation frequency, albeit a function of the amplitude of dynamic displacement.

In Fig. 6, the system damping coefficients ($C_{t_{xx}}, C_{t_{yy}}$) decrease with frequency and also with the amplitude of motion. For the tests with a 13 μm amplitude of motion, the damping is considerably higher at low frequencies. Thus, the system *viscous* damping coefficients show the peculiar features associated with dry-friction energy dissipation. Importantly enough, the effects of the mechanical seal on the system damping coefficients become less pronounced as the frequency and amplitude of motion grow.

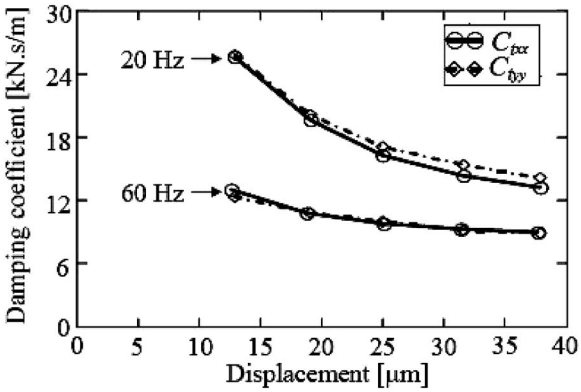


Fig. 8 Identified system damping coefficients ($C_{t_{xx}}, C_{t_{yy}}$) versus displacement amplitude. Unidirectional load tests (excitation frequency: 20 Hz and 60 Hz, lubricated SFD).

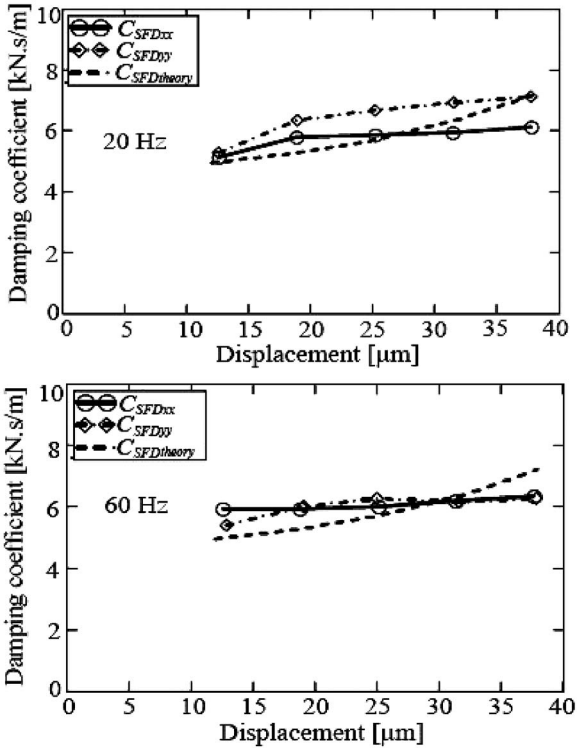


Fig. 9 Squeeze film damping coefficients ($C_{SFD_{xx}}, C_{SFD_{yy}}$) versus displacement amplitude. Unidirectional load tests (excitation frequency: 20 Hz and 60 Hz, lubricated SFD).

Subtracting the "dry friction" equivalent viscous damping (C_{seal}) and remnant structural damping (C_{rv}) from the system damping coefficients, yields the damping coefficients for the squeeze film alone, i.e., $C_{\text{SFD}_{xx}} = C_{t_{xx}} - C_{\text{seal}_x} - C_{\text{rv}_x}$. Figure 7 depicts the SFD coefficients ($C_{\text{SFD}_{xx}}, C_{\text{SFD}_{yy}}$) extracted from the 13 μm and 38 μm dynamic displacement tests. The squeeze film damping coefficients ($C_{\text{SFD}_{xx}}, C_{\text{SFD}_{yy}}$) are nearly constant within the range of test frequencies and similar in magnitude for both displacement amplitudes.

Figure 8 depicts the system total damping coefficients ($C_{t_{xx}}, C_{t_{yy}}$) versus the amplitude of dynamic displacement for excitations at 20 Hz and 60 Hz. Recall that the damper radial clearance is ~0.127 mm, and thus the largest motion amplitudes are up to ~50% of the clearance, i.e., not infinitesimally small. The magnitudes of ($C_{t_{xx}}, C_{t_{yy}}$) are largest for the smallest amplitudes of dynamic motion, in particular at the lowest test frequency, thus evidencing the effect of dry friction interaction upon the dissipation features of the system.

Figure 9 depicts the derived squeeze film damping force coefficients ($C_{\text{SFD}_{xx}}, C_{\text{SFD}_{yy}}$) versus displacement amplitude for the lowest and highest excitation frequencies, 20 Hz and 60 Hz, respectively. The figure includes predictions based on the short length, open ends SFD formulae [4], i.e.

$$C_{\text{theory}} = \eta D \left(\frac{L}{c}\right)^3 \frac{\pi(1+2\varepsilon^2)}{2(1-\varepsilon^2)^{2.5}} \tag{14}$$

The equation above is for a full film (uncavitated) and is derived from small amplitude unidirectional motions about a static journal eccentricity ratio ($\varepsilon = e/c$). The use of the predictive formula for an uncavitated film follows from the verification that there was no lubricant cavitation or air entrainment, as determined from the pressure measurements and flow visualization The test derived squeeze film damping coefficients ($C_{\text{SFD}_{xx}}, C_{\text{SFD}_{yy}}$) increase slightly with the amplitude of motion, though less pro-

nouncedly than the model predictions. The test results at 60 Hz evidence invariant damping coefficients. For both frequencies, the simple model predictions are in good agreement with the identified coefficients. The average value of viscous damping is 6000 N s/m (34.3 lbf s/in.).

Conclusions

The paper describes a test rig and procedure to identify the damping coefficients of a squeeze film damper acting in parallel with a mechanical end seal in contact with the damper journal. The mechanical seal serves to prevent air ingestion and must contain the lubricant within the squeeze film land for extended periods of time. In operation, the mechanical seal also prevents oil leakage from contaminating the ball bearing cartridge while still allowing the damper to provide adequate (designed) dissipation characteristics to reduce undesirable rotor vibrations and mechanical isolation from other components.

The test procedure identifies first the structural stiffness and effective mass of the damper; next the dry-friction force from the mechanical seal, and finally, the test system force coefficients with lubricant flowing through the damper. In the experiments, single frequency (20 Hz–60 Hz) dynamic loads are applied, with magnitudes controlled to maintain prescribed amplitudes of motion. The frequency domain parameter identification builds system transfer functions from measured loads and displacements. The test system damping combines the effects of dry friction in the contact zone of the mechanical seal, the squeeze film lands, and a residual action from the structural supports. The squeeze film damping coefficients are derived from the system damping coefficients after subtraction of the effective viscous damping identified from the dry friction force in the mechanical seal.

The test results show that the system damping coefficients are largest at the lowest excitation frequency and smallest amplitude of motion, denoting the paramount effect of dry friction arising from the mechanical seal. The estimated damping coefficients for the squeeze film land alone are nearly independent of frequency, increasing slightly with the amplitude of dynamic motion. Predictions, based on the short length bearing model, agree well with the experimentally derived squeeze film damping and inertia force coefficients. Importantly enough, the effects of the mechanical seal on the system damping coefficients become less pronounced as the frequency and amplitude of motion increase.

The experimental results demonstrate the mechanical end seal effectively prevents oil leakage and air ingestion into the squeeze film land for the frequencies and amplitudes of motion tested. Future work will include tests with higher frequency ranges providing benchmarks to validate analytical developments to improve predictions of inertia force coefficients in squeeze film dampers.

Acknowledgment

The support from the TAMU Turbomachinery Research Consortium is gratefully acknowledged. Thanks to Mr. Michael Peterson from Honeywell Engines, Systems, and Services (Phoenix) for his technical support and guidance in the design and construction of the test rig.

Nomenclature

c = damper radial clearance (m)

C_{rv_α} = structure remnant damping coefficient $\alpha=x,y$ (N s/m)

C_{seal_α} = equivalent viscous damping coefficient for mechanical seal $\alpha=x,y$ (N s/m)

$C_{t_{\alpha\alpha}}$ = $C_{\mathrm{rv}_\alpha}+C_{\mathrm{seal}_\alpha}+C_{\mathrm{SFD}_{\alpha\alpha}}$ test system damping coefficients $\alpha=x,y$ (N s/m)

$C_{\mathrm{SED}_{\alpha,\beta}}$ = identified squeeze film damping coefficients $\alpha,\beta=x,y$ (N s/m)

E_{dry} = energy dissipated by dry-friction forces (J)

$F_{\alpha(t)}$ = $\bar{F}_\alpha e^{i\omega t}$ external (shaker) forces applied to bearing $\alpha=x,y$ (N)

F_μ = dry friction force from contact in mechanical seal (N)

f_n = natural frequency (Hz)

K_{s_x}, K_{s_y} = structural (support) stiffnesses (N/m)

L, D = film length and diameter of SFD land (m)

M = effective mass of test bearing (kg)

M_f = mass of lubricant in feed plenum and recirculation annulus) (kg)

$M_{\mathrm{SFD}_{\alpha\alpha}}$ = identified squeeze film added mass coefficients $\alpha=x,y$ (kg)

$M_{t_{\alpha\alpha}}$ = $M+M_f+M_{\mathrm{SFD}_{\alpha\alpha}}$ effective mass of lubricated test system $\alpha=x,y$ (kg)

t = time (s)

T = $2\pi/\omega$. period of motion (s)

v = $(\dot{x},\dot{y})$ velocity vector (m/s)

W = work from external forces (J)

X, Y = Cartesian coordinates (m)

x, y = $\bar{x}e^{i\omega t}$, $\bar{y}e^{i\omega t}$ bearing dynamic motions along X, Y directions (m)

ϕ = phase lag of the motion response relative to the excitation force

μ = dry friction coefficient

ρ, η = lubricant density (kg/m^3) and viscosity (Pa s)

ω = excitation frequency (rad/s)

ζ = viscous damping ratio $C/[2(K_s M)^{1/2}]$.

References

[1] Zeidan, F. Y., San Andrés, L., and Vance, J. M., 1996, "Design and Application of Squeeze Film Dampers in Rotating Machinery," *Proc. 25th Turbomachinery Symposium*, Houston, TX, September 17–19, pp. 169–188.

[2] Della Pietra, L., and Adiletta, G., 2002, "The Squeeze Film Damper over Four Decades of Investigations. Part I: Characteristics and Operating Features," Shock Vib. Dig., **34**(1), pp. 3–26.

[3] Della Pietra, L., and Adiletta, G., 2002, "The Squeeze Film Damper over Four Decades of Investigations. Part II: Rotordynamic Analyses with Rigid and Flexible Rotors," Shock Vib. Dig., **34**(2), pp. 97–126.

[4] San Andrés, L., 2006, "SFD Design and Operation," Modern Lubrication Course Notes No. 12, http://phn.tamu.edu/TRIBGroup

[5] Levesley, M., and Holmes, R., 1996, "The Effect of Oil Supply and Sealing Arrangements on the Performance of Squeeze-Film Dampers: an Experimental Study," Proc. Inst. Mech. Eng., Part J: J. Eng. Tribol., **210**(4), pp. 221–232.

[6] De Santiago, O., and San Andrés, L., 1999, "Imbalance Response and Damping Force Coefficients of a Rotor Supported on End Sealed Integral Squeeze Film Dampers," ASME Paper No. 99-GT-203.

[7] Arauz, G., and San Andrés, L., 1997, "Experimental Force Response of a Grooved Squeeze Film Damper," Tribol. Int., **30**(1), pp. 77–86.

[8] San Andrés, L., and Vance, J., 1986, "Experimental Measurements of the Dynamic Pressure Distribution in a Squeeze-Film Bearing Damper Executing Circular-Centered Orbit," ASLE Trans., **30**(3), pp. 373–383.

[9] Kim, K., and Lee, C., 2005, "Dynamic Characteristics of Sealed Squeeze Film Damper With a Central Feeding Groove," ASME J. Tribol., **127**(1), pp. 103–111.

[10] San Andrés, L., and De Santiago, O., 2004, "Forced Response of a Squeeze Film Damper and Identification of Force Coefficients from Large Orbital Motions," ASME J. Tribol., **126**(2), pp. 292–300.

[11] Diaz, S., and San Andrés, L., 2001, "A Model for Squeeze Film Dampers Operating with Air Entrainment and Validation with Experiments," ASME J. Tribol., **123**(1), pp. 125–133.

[12] San Andrés, L., Diaz, S., and Rodriguez, L., 2001, "Sine Sweep Load Versus Impact Excitations and their Influence on the Damping Coefficients of a Bubbly Oil Squeeze Film Damper," Tribol. Trans., **44**(4), pp. 692–698.

[13] Delgado, A., 2005, "Identification of Force Coefficients in a Squeeze Film Damper with a Mechanical Seal," Master of Science thesis, Texas A&M University, College Station TX.

[14] Ginsberg, J. H., 2001, *Mechanical and Structural Vibrations*, Wiley, New York, pp. 135–139.

[15] Beardmore, R., 2005, "Friction Factors," http://www.roymech.co.uk/Useful_Tables/Tribology/co_of_frict.htm

[16] Delgado, A., and San Andrés, L., 2005, "Identification of Structural Stiffness and Damping Coefficients of a Shoed Brush Seal," ASME Paper No. DETC 2005–84159.

[17] San Andrés, L., 2005, "Fluid Inertia in Thin Film Flows," personal communication—Research Notes, Tribology Group, Texas A&M University, Turbomachinery Laboratory, College Station, TX.

Dynamic Properties of Tilting-Pad Journal Bearings: Experimental and Theoretical Investigation of Frequency Effects due to Pivot Flexibility

Waldemar Dmochowski
Institute for Aerospace Research,
Gas Turbine Laboratory,
National Research Council,
Building M-17, 1200 Montreal Road,
Ottawa, Ontario Canada K1A 0R6
e-mail: waldek.dmochowski@nrc-cnrc.gc.ca

Tilting-pad journal bearings (TPJBs) dominate as rotor supports in high-speed rotating machinery. The paper analyzes frequency effects on the TPJB's stiffness and damping characteristics based on experimental and theoretical investigations. The experimental investigation has been carried out on a five pad tilting-pad journal bearing of 98 mm *in diameter. Time domain and multifrequency excitation has been used to evaluate the dynamic coefficients. The calculated results have been obtained from a three-dimensional computer model of TPJB, which accounts for thermal effects, turbulent oil flow, and elastic effects, including that of pad flexibility. The analyzes of the TPJB's stiffness and damping properties showed that the frequency effects on the bearing dynamic properties depend on the operating conditions and bearing design. It has been concluded that the pad inertia and pivot flexibility are behind the variations of the stiffness and damping properties with frequency of excitation.* [DOI: 10.1115/1.2436574]

Keywords: tilting-pad journal bearing, stiffness and damping coefficients

Introduction

Dynamics of high speed rotating machinery depend strongly on journal bearings. Currently, tilting pad journal bearings (TPJBs) are dominant as shaft support in such machinery. This is mainly because of the following two characteristics of the TPJBs: (1) freedom from self-excited vibration; and (2) tolerance to misalignment.

In most practical cases rotordynamic analysis is based on predicted bearing linear dynamic coefficients determined assuming synchronous shaft vibrations (synchronously reduced stiffness and damping coefficients). However, some experimental and theoretical studies of tilting pad journal bearings indicated a certain effect of excitation frequency on the bearing stiffness and damping coefficients [1,2]. Other reported results show rather limited dependency on frequency, or are inconclusive [3]. If the bearing dynamic properties do depend on destabilizing frequency, it is important to also know the properties, which correspond with the first natural frequency. In addition, nonsynchronous forces, such as those associated with internal flow, or due to magnetic effects in generators, may also affect rotor-bearing dynamics.

Pivot stiffness has also been considered an important factor in the bearing dynamic performance [4,5]. Santos [6,7] presented theoretical and experimental studies on controlling the stiffness and damping properties of tilting-pad journal bearings by modifying stiffness of a flexible membrane supporting the pivot.

Obtaining a reliable estimate of the journal bearing dynamic properties has always been a challenging task. The literature documents relatively few cases of such attempts. Using external synchronous loads to excite the test bearing, Glienicke [8] evaluated stiffness and damping properties after measuring the applied load and the resulting shaft orbit. A similar technique was used by Morton [9], who used nonsynchronous excitation in both the vertical and horizontal directions. Parkins [10] and Brockwell and Dmochowski [11] used a more direct method of measuring the bearing dynamic properties of journal bearings by generating two distinct, straight line orbits of vibration. To obtain such orbits, the magnitude of two oscillating forces, as well as the phase difference between them, were carefully adjusted. The main advantage associated with this method was that it avoided the measurement of the phase angle between the vectors of shaft displacement and excitation force.

These methods represent time domain techniques. Unfortunately, it was found that a small error in triggering or in the measured phase angle could result in large errors in the calculated values. Frequency domain techniques overcome these problems. Burrows and Sahinkaya [12], Santos [6,7], and Rodriguez and Childs [1] used frequency domain algorithms for dynamic testing of bearings.

This paper analyzes the variations of the stiffness and damping coefficients for the tilting pad journal bearings with the frequency of excitation and describes the analytical and experimental techniques used to evaluate these properties.

Experimental Investigation

The National Research Council (NRC's) test rig, which is shown in Fig. 1, utilizes the concept of a fixed rotating shaft (1) and a free vibrating test bearing (2). Two orthogonal electromagnetic shakers (3) and (4) apply dynamic loads to the stator, and the bearing's response is measured. Each shaker is attached to the bearing housing through a long steel rod (5). The shaft is supported on high precision, angular ball bearings. A tensioned cable (6) applies a static load. Soft springs (7) minimize the effect of bearing vibration on the applied static load (see Table 1).

The shakers have been programmed to provide a multifrequency excitation, which consisted of frequencies ranging between 20 Hz and 500 Hz with an increment of 10 Hz. In the pres-

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received July 21, 2006; final manuscript received September 1, 2006. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2006: Land, Sea and Air (GT2006), Barcelona, Spain, May 8–11, 2006, Paper No. GT2006-90280.

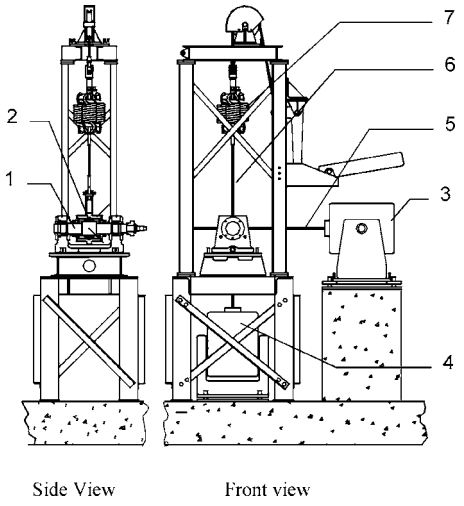


Fig. 1 Journal bearing dynamic test rig

ence of external excitation, equations of bearing motion with respect to the shaft are as follows (Fig. 2(*a*) illustrates the coordinate system)

$$m_b\ddot{x}(t) + k_{xx}x(t) + k_{xy}y(t) + c_{xx}\dot{x}(t) + c_{xy}\dot{y}(t) = f_{d,x}(t)$$

$$m_b\ddot{y}(t) + k_{yx}x(t) + k_{yy}y(t) + c_{yx}\dot{x}(t) + c_{yy}\dot{y}(t) = f_{d,y}(t) \quad (1)$$

The measurement of the bearing/shaft displacement within the bearing clearance is corrupted by noise. Collecting a certain number of records, and using their average can minimize this effect. However, averaging in time domain is not appropriate for random signals, and frequency domain techniques should be applied. Evaluation of the bearing dynamic properties using frequency domain techniques allows for minimization of noise effects and the errors associated with triggering.

After introducing Fourier transforms and power spectral densities, Eq. (1) become

$$G_{F_yF_x}(\omega) - m_bG_{F_yA_x}(\omega) = H_{xx}(\omega)G_{F_yX}(\omega) + H_{xy}(\omega)G_{F_yY}(\omega)$$

Table 1 Rig specifications

Shaft speed	16,500 rpm
Journal diameter	0.09843 m (3.875)
Static load	20 kN (4500 lbf)
Lubricant flow	0.44 m^3/s (7 US gpm)
Lubricant inlet temperature	70°C (158°F),
Electric motor (variable)	37 kW (50 HP)
System's natural frequency	Approx. 120 Hz

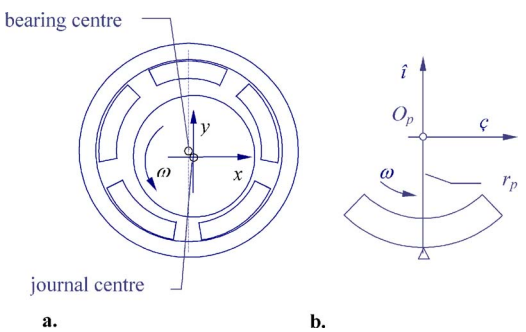


Fig. 2 Coordinate systems: (*a*) attached to shaft center; and (*b*) attached to pad center

$$G_{F_yF_y}(\omega) - m_bG_{F_yA_y}(\omega) = H_{yx}(\omega)G_{F_yX}(\omega) + H_{yy}(\omega)G_{F_yY}(\omega) \quad (2)$$

where H_{ij} is the frequency response function

$$H_{ij} = k_{ij} + i\omega c_{ij} \quad (3)$$

The power spectral densities $G_{uu}(\omega)$ (autospectral density) and $G_{uv}(\omega)$ (cross spectral density) are defined as

$$G_{uv}(\omega) = \lim_{T\to\infty}\frac{2}{T}E[u^*(\omega)v(\omega)] \quad (4)$$

and * denotes complex conjugate.

In order to calculate the power spectral densities from the test data one can use the following formula [13]

$$G_{uv}(\omega) \approx \lim_{N\to\infty}\frac{2}{n_dN\Delta t}\sum_{i=1}^{n_d}u_i^*(\omega)v_i(\omega) \quad (5)$$

The two Eq. (2) contain four unknowns H_{ij}. Thus, two independent excitations are required to evaluate the bearing dynamic properties. These properties have been determined from 32 consecutive records. Each record consisted of 512 samples of each measured variable (six channels), collected over a period of 0.1 s. Before sampling an anti-aliasing filtering with a cutoff frequency of 500 Hz has been used.

Computer Model

Calculations of the bearing dynamic properties have been based on a three-dimensional model of tilting-pad journal bearings, though all the major equations are reduced to a two-dimensional form. The model has been described elsewhere [14]. Here, only a brief description is given.

A finite length pressure equation (circumferential and axial directions) allows for viscosity variations in circumferential direction and across the oil film. Turbulent flow is also accounted for by including Reynolds number effect. The temperature and viscosity fields are obtained from a two-dimensional energy equation, which accounts for heat conduction in radial direction and heat convection circumferentially. Oil mixing in the bearing cavities as well as hot-oil carryover is also included in the analysis. Heat conducted through the pad is calculated from the Laplace equation, which accounts for heat conduction in circumferential and radial directions. The model also calculates both the thermal and elastic distortions of the individual pads.

The above hydrodynamic considerations allow for calculation of the stiffness and damping coefficients for fixed pads. They have been used to calculate the coefficients for individual tilting pads applying the technique described by Lund [15]. This technique takes into considerations the mass and the excitation frequency of the tilting pad. For example, in the coordinate system associated with the pad (Fig. 2(*b*)), the direct coefficients of stiffness and damping in radial direction for the "*i*" pad can be written as

$$k_{\xi\xi}^{t(i)} = k_{\xi\xi}^{(i)} + \frac{1}{A_\eta^{(i)}}[(k_{\eta\eta}^{(i)} + m_z^{(i)}\omega_{\text{exc}}^2)(\omega_{\text{exc}}^2c_{\xi\eta}^{(i)}c_{\eta\xi}^{(i)} - k_{\xi\eta}^{(i)}k_{\eta\xi}^{(i)}) - \omega_{\text{exc}}^2c_{\eta\eta}^{(i)}(k_{\xi\eta}^{(i)}c_{\eta\xi}^{(i)} + k_{\eta\xi}^{(i)}c_{\xi\eta}^{(i)})] \quad (6)$$

$$c_{\xi\xi}^{t(i)} = c_{\xi\xi}^{(i)} - \frac{1}{A_\eta^{(i)}}[(k_{\eta\eta}^{(i)} + m_z^{(i)}\omega_{\text{exc}}^2)(k_{\xi\eta}^{(i)}c_{\eta\xi}^{(i)} + k_{\eta\xi}^{(i)}c_{\xi\eta}^{(i)}) + c_{\eta\eta}^{(i)}(\omega_{\text{exc}}^2c_{\xi\eta}^{(i)}c_{\eta\xi}^{(i)} - k_{\xi\eta}^{(i)}k_{\eta\xi}^{(i)})] \quad (7)$$

where

$$A_\eta^{(i)} = (k_{\eta\eta}^{(i)} + m_z^{(i)}\omega_{\text{exc}}^2)^2 + \omega_{\text{exc}}^2c_{\eta\eta}^{(i)} \quad (8)$$

$$m_z^{(i)} = \frac{I_p^{(i)}}{r_p^{(i)}} \quad (9)$$

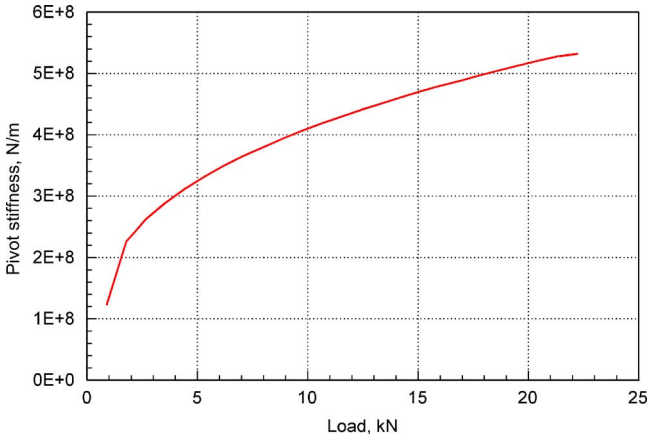


Fig. 3 Typical pivot stiffness of a 100 mm TPJB

Relationships for the remaining six coefficients are given in the Appendix.

Typically, calculated bearing dynamic properties are obtained after assuming that pad mass can be neglected and that both shaft and pad motions are synchronous with shaft speed. Indeed, mass forces associated with tilting of the pads are negligible when compared with viscous forces. However, pivot stiffness can be of the same order of magnitude as the oil film stiffness, and thus can play an important role in the bearing dynamic properties. Figure 3 shows pivot stiffness for a typical 100-mm-diameter bearing calculated using the formulae given by Kirk and Reedy [16]. Thus, each pad can be represented by the mass, spring, and damper elements, as shown in Fig. 4.

The shaft experiences the combined action of all the elements of the system represented by equivalent stiffness and damping coefficients, which can be evaluated from the following formulae

$$k_{ij}^{(\mathrm{eq})} = \frac{k_{ij}k_z(k_{ij}+k_z) - \omega_{\mathrm{exc}}^2 k_z c_{ij}^2}{(k_{ij}+k_z)^2 + \omega_{\mathrm{exc}}^2 c_{ij}^2} \tag{10}$$

$$c_{ij}^{(\mathrm{eq})} = \frac{k_z^2 c_{ij}}{(k_{ij}+k_z)^2 + \omega_{\mathrm{exc}}^2 c_{ij}^2} \tag{11}$$

where

$$k_z = m_p\omega_{\mathrm{exc}}^2 + k_{\mathrm{piv}} \tag{12}$$

i,j denote directions ξ or η. Vector summations of the $k_{i,j}^{(i)}$ and $c_{i,j}^{(i)}$ for all the pads give the bearing dynamic coefficients.

Results and Discussion

In order to evaluate the frequency effects on the bearing dynamic properties calculations, an experimental investigation has been carried out for two operating conditions, which are typical for high speed rotating machinery. The bearing parameters and operating conditions are shown in Table 2. For both cases pivot stiffness is similar and its variation with load is shown in Fig. 3.

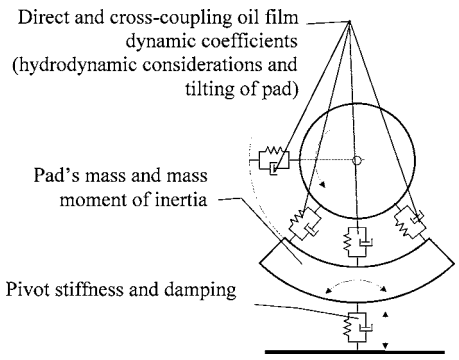


Fig. 4 Mass, spring, and damper elements for tilting pad

Table 2 Bearing parameters and test conditions

	Case 1	Case 2
Bearing type	5 pad TPJB	5 pad TPJB
Pad	Load between pads	Load on pad
Nominal	99 mm (3.88 in.)	100 mm (4.0 in.)
L/d ratio	0.4	1.0
Assembly	0.09 mm (0.0035 in.)	0.09 mm (0.0035 in.)
Preload	0.3	0.3
Bearing load	4.03 kN (900 lbf.)	4.45 kN (1000 lbf.)
Shaft speed	8900 rpm	9000 rpm

Suitability of the test arrangement and expected quality of the results has been confirmed by the evaluated coherence functions. As an example, Fig. 5 illustrates the coherence of the vertical excitation and displacement for Case 1 (Table 2).

Case 1 represents a moderately loaded bearing. Figure 5 illustrates the measured and calculated variations of the real part of the frequency response function, Eq. (3), which represents bearing stiffness. The results show a certain frequency effect on the bearing stiffness properties; the coefficients decrease with the excitation frequency. For the excitation frequencies up to that of the shaft rotation (150 Hz), coefficients of stiffness are relatively constant. This effect is particularly clear for the vertical bearing stiffness, $k_{yy}=\mathrm{Re}(H_{yy})$, at higher frequencies of excitation.

The direct bearing damping coefficients for Case 1 are represented by the slope of the imaginary part of the frequency response function $H_{i,j}$ (Eq. (3)) illustrated in Fig. 6. Constant slope indicates negligible effects of both the pivot flexibility and the pad's mass.

Although the trends of the experimental results shown in Figs. 6 and 7 are well defined, a certain spread of the results can be

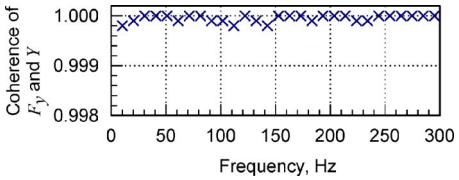


Fig. 5 Coherence function F_y and Y

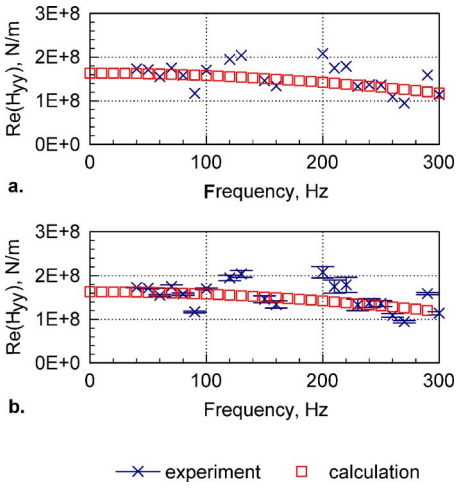


Fig. 6 Bearing direct stiffness coefficients: Case 1: (*a*) horizontal; and (*b*) vertical

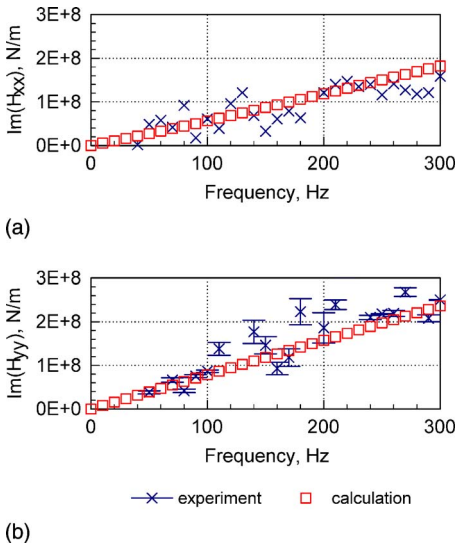


Fig. 7 Bearing direct damping properties: Case 1: (*a*) horizontal; (*b*) vertical

seen. The power spectral density method offers improved accuracy of the bearing dynamic measurement. Typically, uncertainty of the measured bearing dynamic properties is evaluated based on statistical analysis of the obtained data, which is usually referred to as Type A analysis [17] and gives an idea about the consistency of the results. The analysis described in Ref. [18] estimates the effect of elemental uncertainties of individual measurements on the uncertainties of the frequency response functions. Such an approach is referred to as Type B analysis [17]. The uncertainties shown in Figs. 5 and 6 did not exceed 12% either for the real or imaginary part of the frequency response function. The elemental uncertainties were 2.5 μm (0.0001 in.) for displacement and 0.01 g for acceleration.

With the validated model from Case 1, a different bearing dynamic behavior has been observed in Case 2 (Table 2), which deals with similar bearing operating conditions and bearing geometry as those of Case 1, except for the bearing width. The same pivot stiffness (Fig. 3) has been used in the calculations. With an L/d ratio of 1.0, the bearing is considered to be lightly loaded, as the operating eccentricity was approximately 0.1 of the bearing radial clearance. At such a low eccentricity, and with the preload of 0.3, each pad of the bearing operates under load. In this situation, both the horizontal and vertical bearing direct stiffness coefficients increase with the frequency of excitation (Fig. 8).

As opposed to Case 1, frequency of excitation very strongly affected the bearing damping properties. Figure 9 shows that the imaginary part of the frequency response function levels out, which means a significant decrease in bearing damping, even at subsynchronous (with respect to the shaft rotation) frequencies.

Figures 8 and 9 also compare the effect of pad inertia in tilting and radial motion (enabled by pivot flexibility). It is the pad radial movement and pivot stiffness that lead to variations of the dynamic properties with frequency of excitation. Bearings with rigid pivots have constant stiffness and damping coefficients for the entire range of considered frequencies.

Figures 8(*b*) and 9(*b*) show that the pad flexibility and associated inertia forces do not affect the cross-coupling coefficients for tilting-pad journal bearings.

Concluding Remarks

Pivot flexibility can have a significant effect on the TPJB's dynamic properties, in particular at higher frequencies of excitations.

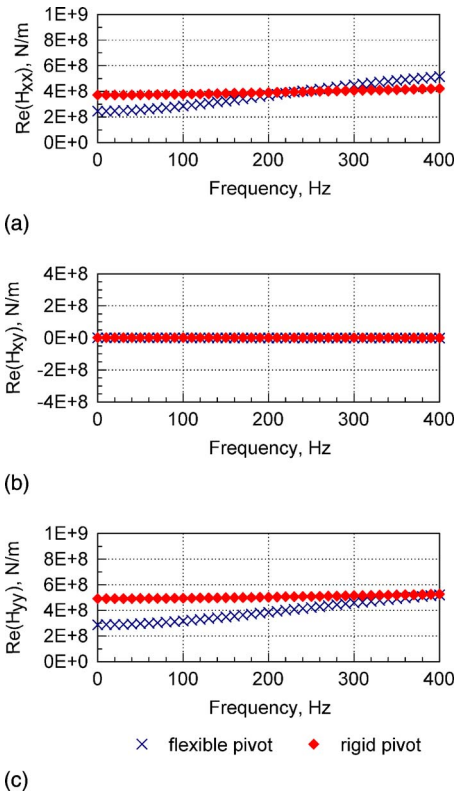


Fig. 8 Calculated bearing stiffness coefficients: Case 2: (*a*) horizontal; (*b*) cross-coupling; and (*c*) vertical

As a result of pivot flexibility, the bearing stiffness coefficients can increase or decrease with the frequency of excitation, depending on operating conditions and bearing design.

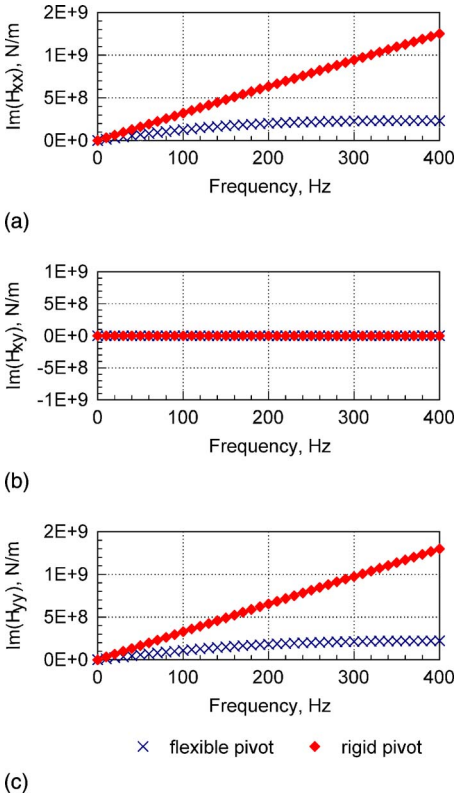


Fig. 9 Calculated bearing damping properties: Case 2: (*a*) horizontal; (*b*) cross-coupling; and (*c*) vertical

In the presence of pivot flexibility, an increase in frequency of excitation can lead to a significant decrease in bearing damping in the horizontal as well as in the vertical directions.

Acknowledgment

The author wishes to thank Waukesha Bearing Corporation for permission to publish the experimental results.

Nomenclature

A_x, A_y = Fourier transforms of accelerations in the horizontal and vertical directions, respectively

F_x, F_y = Fourier transforms of excitation forces in the horizontal and vertical directions, respectively

I_p = mass moment of inertia of pad about pivot

L = bearing width

N = number of data points

T = sampling time

X, Y = Fourier transforms of displacements in the horizontal and vertical directions, respectively

$c_{xx}, c_{xy}, c_{yx}, c_{yy}$ = bearing damping coefficients in x-y coordinate system (see Fig. 2(a))

$c_{\xi\xi}^{(\mathrm{eq})}, c_{\xi\eta}^{(\mathrm{eq})}, c_{\eta\xi}^{(\mathrm{eq})}, c_{\eta\eta}^{(\mathrm{eq})}$ = equivalent pad damping coefficients in ξ-η coordinate system (see Fig. 2(b))

$c_{\xi\xi}^{(i)}, c_{\xi\eta}^{(i)}, c_{\eta\xi}^{(i)}, c_{\eta\eta}^{(i)}$ = fixed pad damping coefficients in ξ-η coordinate system (see Fig. 2(b))

$c_{\xi\xi}^{t(i)}, c_{\xi\eta}^{t(i)}, c_{\eta\xi}^{t(i)}, c_{\eta\eta}^{t(i)}$ = tilting pad damping coefficients in ξ-η coordinate system (see Fig. 2(b))

d = bearing nominal diameter

$f_{d,x}, f_{d,y}$ = components of the dynamic force (excitation)

$k_{xx}, k_{xy}, k_{yx}, k_{yy}$ = bearing stiffness coefficients in x-y coordinate system (see Fig. 2(a))

$k_{\xi\xi}^{(eq)}, k_{\xi\eta}^{(eq)}, k_{\eta\xi}^{(eq)}, k_{\eta\eta}^{(eq)}$ = equivalent pad damping coefficients in ξ-η coordinate system (see Fig. 2(b))

$k_{\xi\xi}^{(i)}, k_{\xi\eta}^{(i)}, k_{\eta\xi}^{(i)}, k_{\eta\eta}^{(i)}$ = fixed pad stiffness coefficients in ξ-η coordinate system (see Fig. 2(b))

$k_{\xi\xi}^{t(i)}, k_{\xi\eta}^{t(i)}, k_{\eta\xi}^{t(i)}, k_{\eta\eta}^{t(i)}$ = tilting pad stiffness coefficients in ξ-η coordinate system (see Fig. 2(b))

m_b = bearing mass

m_p = mass of pad

m_z = equivalent pad mass, $m_z = I_p / r_p$

n_d = number of records

r_p = distance from pivot to pad center

x, y = shaft center coordinates in the rectangular system with the origin at the bearing equilibrium position

Δt = time increment for sampling data

ω_{exc} = frequency

ω = angular speed of the shaft

Appendix

Equations (6) and (7) describe the radial direct stiffness and damping coefficients for an individual "i" tilting pad. The remaining six dynamic coefficients are as follows

$$k_{\xi\eta}^{t(i)} = \frac{1}{A_\eta^{(i)}}[(k_{\eta\eta}^{(i)} + m_z^{(i)}\omega_j^2)k_{\xi\eta}^{(i)} + \omega_j^2 c_{\eta\eta}^{(i)} c_{\xi\eta}^{(i)}]m_z^{(i)}\omega_{\mathrm{exc}}^2 \quad \text{(A1)}$$

$$k_{\eta\xi}^{t(i)} = \frac{1}{A_\eta^{(i)}}[(k_{\eta\eta}^{(i)} + m_z^{(i)}\omega_{\mathrm{exc}}^2)k_{\eta\xi}^{(i)} + \omega_{\mathrm{exc}}^2 c_{\eta\eta}^{(i)} c_{\eta\xi}^{(i)}]m_z^{(i)}\omega_{\mathrm{exc}}^2 \quad \text{(A2)}$$

$$k_{\eta\eta}^{t(i)} = \left[1 - \frac{1}{A_\eta^{(i)}}(k_{\eta\eta}^{(i)} + m_z^{(i)}\omega_{\mathrm{exc}}^2)m_z^{(i)}\omega_{\mathrm{exc}}^2\right]m_z^{(i)}\omega_{\mathrm{exc}}^2 \quad \text{(A3)}$$

$$c_{\xi\eta}^{t(i)} = \frac{1}{A_\eta^{(i)}}[(k_{\eta\eta}^{(i)} + m_z^{(i)}\omega_{\mathrm{exc}}^2)c_{\xi\eta}^{(i)} + k_{\xi\eta}^{(i)} c_{\eta\eta}^{(i)}]m_z^{(i)}\omega_{\mathrm{exc}}^2 \quad \text{(A4)}$$

$$c_{\eta\xi}^{t(i)} = \frac{1}{A_\eta^{(i)}}[(c_{\eta\eta}^{(i)} + m_z^{(i)}\omega_{\mathrm{exc}}^2)c_{\eta\xi}^{(i)} - k_{\eta\xi}^{(i)} c_{\eta\eta}^{(i)}]m_z^{(i)}\omega_{\mathrm{exc}}^2 \quad \text{(A5)}$$

$$c_{\eta\xi}^{t(i)} = \frac{1}{A_\eta^{(i)}} c_{\eta\eta}^{(i)} (m_z^{(i)}\omega_{\mathrm{exc}}^2)^2 \quad \text{(A6)}$$

where $A_\eta^{(i)}$ and $m_z^{(i)}$ are given by Eqs. (8) and (9).

References

[1] Rodriguez, L. E., and Childs, D. W., 2004, "Experimental Rotordynamic Coefficient Results for a Load-On-Pad Flexible-Pivot Tilting-Pad Bearing with Comparisons to Predictions from Bulk-Flow and Reynolds Equations Models," Proc., 2004 ASME/STLE International Joint Tribology Conference, Long Beach, CA, October 24–27, Paper TRIB 2004–64042.

[2] Parsell, J. K., Allaire, P. E., and Barret, L. E., 1983, "Frequency Effects in Tilting-Pad Journal Bearing Dynamic Coefficients," ASLE Trans. **26**, pp. 222–227.

[3] Ha, H. C., and Yang, S. H., 1999, "Excitation Frequency Effects on the Stiffness and Damping Coefficients of a Five-Pad Tilting Pad Journal Bearing," ASME J. Tribol. **121**(3), pp. 517–522.

[4] Klumpp, R., 1976, "Die Eigenschaften von Kippsegment-Radiallagern," Konstruktion, **28**(8), pp. 320–324.

[5] Springer, H., 1980, "Dynamische Eigenschaften von Gleitlagern mit beweglischen Segmenten," VDI-Ber. **381**, pp. 177–184.

[6] Santos, I. F., 1995, "On the Adjusting of the Dynamic Coefficients of Tilting Pad Bearings," STLE Tribol. Trans. **38**(3), pp. 700–706.

[7] Santos, I. F., 1996, "Theoretical and Experimental Identification of the Stiffnes and Damping Coefficients of Active Tilting-Pad Journal Bearings," *Identification in Engineering Systems*, M. I. Friswell and J. E. Mottershead, eds., The Cromwell Press Ltd., Swansea, pp. 325–334.

[8] Glienicke, J., 1967, "Experimental Investigation of the Stiffness and Damping Coefficients of Turbine Bearings and Their Application to Instability Prediction," Proc. Inst. Mech. Eng., **181**(3B), pp. 116–129.

[9] Morton, P. G., 1971, "Measurement of the Dynamic Characteristics of a Large Sleeve Bearing," ASME J. Lubr. Technol. **93**(1), pp. 143–150.

[10] Parkins, D. W., 1979, "Theoretical and Experimental Determination of the Dynamic Characteristics of a Hydrodynamic Journal Bearing," ASME J. Lubr. Technol. **101** pp. 120–125.

[11] Brockwell, K. R., and Dmochowski, W., 1989, "Experimental Determination of the Journal Bearing Oil Film Coefficients by the Method of Selective Vibration Orbits," *Proc. 12th Biennial ASME Conference on Mechanics Vibration and Noise*, Montreal, September 17–21, DE-Vol. 18–1, pp. 251–259.

[12] Burrows, C. R., and Sahinkaya, M. N., 1982, "Frequency-Domain Estimation of Linearized Oil-Film Coefficients," ASME J. Lubr. Technol. **104**(2), pp. 210—215.

[13] Bendat, J. S., and Piersol, A. G., 1996, *Random Data. Analysis and Measurement Techniques*, Wiley, New York.

[14] Dmochowski, W., Brockwell, K., DeCamillo, S., and Mikula, A., 1993, "A Study of the Thermal Characteristics of the Leading Edge and Conventional Tilting Pad Journal Bearings," ASME J. Tribol. **115**(2), pp. 219–226.

[15] Lund, J. W., 1964, "Spring and Damping Coefficients For the Tilting-Pad Journal Bearing," ASLE Trans. **7**(3), pp. 342–352.

[16] Kirk, R. G., and Reedy, S. W., 1988, "Evaluation of Pivot Stiffness for Typical Tilting-Pad Journal Bearing Designs," ASME J. Vib., Acoust., Stress, Reliab. Des. **110**, pp. 165–171.

[17] International Organization of Standarization, 1993, *Guide to the Expressions of Uncertainty in Measurement*, Geneva, Switzerland.

[18] Dmochowski, W., and Dmochowski, J., 2005, "Frequency Dependent Dynamic Properties Of Tilting Pad Journal Bearings: Experimental Results And Uncertainty Analysis," *Proc. World Tribology Congress III*, Washington, D.C., September 12–16, Paper WTC2005–63274.

Experimental and Computational Studies on Gasoline HCCI Combustion Control Using Injection Strategies

Zhi Wang
e-mail: wangzhi@tsinghua.edu.cn

Jian-Xin Wang
e-mail: wangjx@tsinghua.edu.cn

Shi-Jin Shuai

Guo-Hong Tian

Xin Liang An

State Key Laboratory of Automotive Safety and Energy,
Tsinghua University,
Beijing 100084, China

Homogeneous Charge Compression Ignition (HCCI) has challenges of ignition control. In this paper, HCCI ignition timing and combustion rate were controlled by two-stage direct injection (TSDI) strategies on a four-stroke gasoline HCCI engine. TSDI strategy was proposed to solve the two major problems of HCCI application-ignition control and load extension. Both simulation and experiments were carried out on a gasoline HCCI engine with negative valve overlap (NVO). An engine model with detailed chemical kinetics was established to study the gas exchange process and the direct injection strategy in the gasoline HCCI engine with TSDI and NVO. Simulation results were compared with experiments and good agreement was achieved. The simulated and experimental results provided a detailed insight into the processes governing ignition in the HCCI engine. Using TSDI, the fuel concentration, temperature as well as chemical species can be controlled. The effects of different injection parameters, such as split injection ratio and start-of-injection (SOI) timing, were studied. The experimental results indicate that, two-stage direct injection is a practical technology to control the ignition timing and combustion rate effectively in four-stroke gasoline HCCI engines. Both the high load and low load limits of HCCI operation were extended. [DOI: 10.1115/1.2227030]

Introduction

Homogeneous Charge Compression Ignition (HCCI) combustion has advantages of high thermal efficiency and low emissions and is becoming a promising combustion mode in internal combustion (IC) engines. However, HCCI still faces the challenges of ignition timing control, combustion rate control, and operating range extension [1].

It is generally agreed that HCCI combustion is a chemical kinetics controlled combustion [2]. This view has been supported by spectroscopic data which revealed that the order of radical formation in HCCI combustion corresponds to self-ignition rather than flame propagation [3,4]. Therefore, detailed chemical kinetics must be taken into account for HCCI combustion modeling.

Single-zone or multizone models [5,6] require initial conditions to be specified at the IVC, including the average temperature, pressure of the mixture and each concentration of species in the cylinder. These initial conditions are difficult to obtain in a test engine. Considering the importance of these initial conditions for simulation accuracy, 1D HCCI engine models combining both detailed combustion chemistry and gas exchange processes need to be developed so that the initial conditions at IVC no longer have to be set. Generally, HCCI engine cycle models are constructed by combining the CHEMKIN code package and 1D engine simulation code. The gas exchange process affects the engine parameters and charge properties and therefore plays a significant role in determining the control of the HCCI process [7]. Therefore, 1D modeling is usually used to analyze the influence of the variable valve timing strategy on the gas exchange process [8].

In HCCI combustion, since there are no compulsive triggers, such as spark in traditional gasoline engines and fuel injection in diesel engines, optimization of combustion phase is more difficult than that in SI or diesel engines. HCCI combustion is controlled by chemical kinetics. The ignition timing and combustion rate in the HCCI engine can be adjusted by intake temperature, boost pressure, compression ratio, EGR ratio, and fuel property. In all these parameters, the temperature is the most sensitive one, but air temperature control usually has a slow response. Meanwhile, with the engine speed and load change, the ignition timing varies as well. To control HCCI combustion, several indirect control methods which influence HCCI combustion have been applied to gasoline HCCI investigation, including air preheating [9], supercharging [10], VCR [11], EGR [12], etc. In order to control ignition timing, additional devices are needed, for instance, dual-fuel with different ignition property [13], mixing peroxide additive to improving ignition property [14], and introducing a laser to stimulate HCCI ignition [15]. But these methods or devices are not suitable for practical application on a commercial engine currently.

Therefore, stable HCCI combustion can be achieved at different operation conditions and HCCI engines can become commercial products only if a robust method with fast response and easy adjustment in engine operation are found.

According to the Arrhenius formula, $k_{fi}=A_i T^{\beta_i} \exp[-(E_i/RT)]$, the three factors which influence combustion are temperature, concentration, and components. For temperature, it has a slow response to control. However, it is possible to control the concentration and components precisely and rapidly. After the intake valve closure, there are three methods to influence the autoignition process. The first one is achieving the inhomogeneity of temperature and concentration distribution using direct fuel injection, the second one is introducing external energy such as laser stimulation or high-energy spark ignition, and the third one is fuel reforming to enhance the ignition property.

This paper presents an injection strategy using two-stage direct injection for compression ignition and solves the problems of ignition control and extending the operation range. The objective of the work is to reveal the ignition and combustion characteristics in direct injection compression ignition engines.

Contributed by the Internal Combustion Engine Division of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received November 2, 2005; final manuscript received February 21, 2006. Review conducted by J. Cowart.

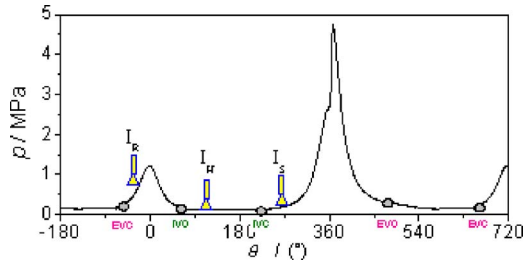


Fig. 1 Injection strategies of HCCI engines with NVO

NVO-TSDI Concept

Negative valve overlap (NVO) is a practical approach to obtain internal EGR. Figure 1 shows the typical indicated diagram of a four-stroke gasoline HCCI engine with NVO. Intake TDC and combustion TDC are defined as 0 deg and 360 deg, respectively, in this paper. The second peak of cylinder pressure and temperature occurs at intake TDC due to NVO. This makes it possible to utilize thermal atmosphere for fuel evaporation and reforming. When it combines with multiple direct injection strategies, three CI combustion modes, i.e., Reformed Charge Compression Ignition (RCCI), HCCI, and Stratified Charge Compression Ignition (SCCI) can be achieved.

RCCI Mode. During the period of NVO, part of the exhaust gas is trapped in the cylinder and recompressed. During the NVO period the direct injected fuel (I_R) is reformed due to the high temperature and pressure of internal EGR, and an amount of active species present, which improve the ignition property [16]. This combustion mode is called Reformed Charge Compression Ignition (RCCI) in this paper. RCCI helps to improve the combustion stability, extend the operation range at low load, and decrease the HC and CO emissions. The most important thing is that the ignition timing can be controlled by I_R.

HCCI Mode. Early fuel injection in the intake stroke (I_H) can obtain HCCI due to sufficient mixing time available to form a homogeneous charge [17,18]. The engine load can be controlled by I_H.

SCCI Mode. The injection in the compression process (I_S) can form a stratified charge. Stratified Charge Compression Ignition (SCCI) combustion has capabilities to suppress knocking [19], reduce HC emission, extend the high load limit, and is suitable for transient operation.

Combining the above three injection strategies, and adjusting fuel quantity and injection timing of each pulse, the temperature, concentration, and species of the mixture in the cylinder can be controlled flexibly. As a result, the HCCI ignition timing and combustion rate can be controlled.

Model Description

In this paper, 1D engine cycle simulation code (AVL BOOST) with detailed combustion chemistry is applied to simulate the HCCI engine working process and used to optimize the direct injection strategies in a gasoline HCCI engine with negative valve overlap.

Combustion. A single-zone High-RON Gasoline (HRG) combustion model, which accounts for valve timing, heat transfer, blow-by, and detailed reaction mechanism (89 species, 413 reactions) [20], was developed based on the original CHEMKIN II code. In this model, the cylinder is regarded as a variable-volume fixed-mass reactor. The reaction mechanism was implemented into the single-zone model to evaluate the HCCI process during the compression and expansion strokes (from IVC to EVO).

The relationship between the total volume available for combustion and the clearance volume is given as a function of crank angle

$$\frac{V}{V_c} = 1 + \frac{1}{2}(\varepsilon - 1)[\lambda + 1 - \cos\theta - (\lambda^2 - \sin^2\theta)^{1/2}]$$

where θ is the crank angle degree, ε is the compression ratio, V_c is the clearance volume, and λ is the ratio of the connecting rod length to the crank-arm radius.

Heat Transfer. In the HCCI engine modeling, heat loss due to heat transfer from cylinder walls cannot be neglected. Considering the heat transfer, the energy conservation equation was modified in this paper,

$$C_v\frac{dT}{dt} + p\frac{dv}{dt} + v\sum_{k=1}^{K} e_k\dot{\omega}_k W_k = \frac{dq}{dt}$$

where C_v is mean specific heat of the mixture. T is temperature, p is pressure, v is specific volume, e_k is internal energy per mass, $\dot{\omega}_k$ is the molar production rate of the kth species, W_k is the molecular weight of the kth species, and q is heat transfer per mass.

Considering the homogeneity and rapid heat release characteristic of HCCI combustion, the Woschni heat transfer model which is widely applied in traditional diesel engine, is not suitable. In this paper, the Hohenberg heat transfer model [21] which is suitable for low temperature and low load operation was adopted,

$$\frac{dq}{dt} = hA(T - T_{\text{wall}})$$

$$h = 105 \cdot V^{-0.06} \cdot p^{0.8} \cdot T^{-0.4} \cdot (\overline{S_p} + 1.4)^{0.8}$$

where h is heat transfer coefficient, A is the total area of cylinder walls, $\overline{S_p}$ is mean piston velocity.

Blow-by. Blow-by losses in the cylinder is considered using specified effective blow-by gap and mean crankcase pressure. The blow-by mass flow rates are calculated at any time step from Bernoulli equations

$$\dot{m} = \rho A_b\sqrt{\frac{2\Delta p}{\rho}}$$

The effective flow area is calculated from cylinder bore diameter and effective blow-by gap

$$A_b = \pi D\delta$$

where A_b is effective flow area, D is cylinder bore diameter, δ is blow-by gap.

Gas Exchange Process. When the valves open, the equation for the gas exchange process is based on the first law of thermodynamics

$$\frac{d(m_c \cdot u)}{d\theta} = -p_c\frac{dV}{d\theta} - \sum\frac{dQ_w}{d\theta} + \sum\frac{dm_i}{d\theta}h_i - \sum\frac{dm_e}{d\theta}h_e$$

where $[d(m_c \cdot u)]/d\theta$ is the change of internal energy, $-p_c(dV/d\theta)$ is the work of piston, $-\Sigma(dQ_w/d\theta)$ is heat transfer from cylinder wall, p_c is cylinder pressure, dm_i is mass element flowing into the cylinder, dm_e is mass element flowing out of the cylinder, h_i is enthalpy of the in-flowing mass, h_e is enthalpy of the mass leaving the cylinder.

High Pressure Process In-Cylinder. When valves close, the calculation of high pressure cycle is

Table 1 Engine specifications

Type	2-cylinder in-line, four-stroke; DI
Number of valves	2 per cylinder
Bore	95 mm
Stroke	115 mm
Connecting rod	210 mm
Compression ratio	13
Throttle	WOT
EGR	Internal EGR
Intake temperature	Ambient
Fuel	Gasoline (RON93)
Injector	High pressure swirl injector
Injection pressure	5 MPa
Head bottom	Flat
Intake port	Helical

$$\frac{\mathrm{d}(m_c \cdot u)}{\mathrm{d}\theta} = -p_c \frac{\mathrm{d}V}{\mathrm{d}\theta} + \sum \frac{\mathrm{d}Q_F}{\mathrm{d}\theta} - \sum \frac{\mathrm{d}Q_w}{\mathrm{d}\theta} - \frac{\mathrm{d}m_B}{\mathrm{d}\theta} h_B$$

where $\mathrm{d}Q_F/\mathrm{d}\theta$ is the fuel heat input, $(\mathrm{d}m_B/\mathrm{d}\theta)h_B$ is enthalpy flow due to blow-by.

Two-Stage Direct Injection In-Cylinder. Injection timing and split injection quantity are the most important parameters in the HCCI engine with TSDI. According to the approach of Koopmans et al. [22], the additional two-stage timing and fuel mass in-cylinder by direct injection is accounted for in the modified equation

$$C_v \frac{\mathrm{d}T}{\mathrm{d}t} + p\frac{\mathrm{d}v}{\mathrm{d}t} + v\sum_{k=1}^{K} e_k \dot{\omega}_k W_k - \frac{\mathrm{d}q_{\mathrm{inj}}}{\mathrm{d}t} = \frac{\mathrm{d}q_{\mathrm{loss}}}{\mathrm{d}t}$$

where $\dot{m}_{\mathrm{inj},k}$ is the fuel mass input, $\mathrm{d}q_{\mathrm{inj}}/\mathrm{d}t$ is the latent heat of fuel evaporation rate, $\mathrm{d}q_{\mathrm{loss}}/\mathrm{d}t$ is the heat transfer rate.

In this mode, the fuel spray, droplet breakup, and air motion are neglected. The mixture is regarded as homogeneous gas consisting of fuel, air, and EGR. Despite these simplifications, the model is useful in the analysis of an injection strategy with negative valve overlap.

Experimental Setup

The investigation was carried out on a dual-cylinder four-stroke test engine. The engine specifications are listed in Table 1. The properties of the test fuel are listed in Table 2.

The schematic diagram of the combustion system is shown in Fig. 2. The additional spark is used in the spark ignition (SI) combustion mode at high load. High intake swirl can be kept in the toroidal combustion chamber promoting dispersion of the fuel injected in compression stroke.

Two HCCI combustion modes with different cam-phases were obtained by replacing cams. The cam-phase configurations in the experiments are list in Table 3.

HCCI Engine Cycle Simulation

Model Validation. Figure 3 shows the comparison of motored and firing pressure traces between experiments and simulation. It can be seen that the calculated pressure curves match the experiments very well except for the peak cylinder pressure in the HCCI case due to the characteristics of single-zone combustion model. The ignition timing, IMEP, and rate of pressure rise of the simulations show good agreement with the experimental results. Therefore, the established engine cycle mode can be used to parametrically study the HCCI engine performance.

Table 2 Properties of the test fuel

Gasoline	93#
RON	94.4
MON	83.0
Oxygen content (wt.)	0.21
Density (20°C)(g/cm^3)	0.7487
C/H	6.42
Stoichiometric A/F ratio	14.6

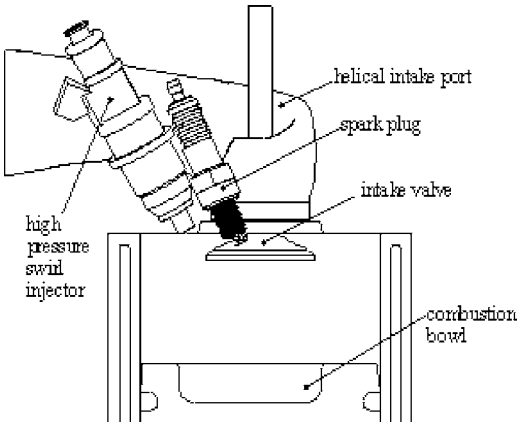


Fig. 2 Combustion system layout

Table 3 Cam-phase configuration (°CA)

	HCCI(1)	HCCI(2)	HCCI(3)
EVO BBDC	44	52	45
EVC ATDC	−58	−66	−75
IVO BTDC	−64	−80	−75
IVC ABDC	24	40	45

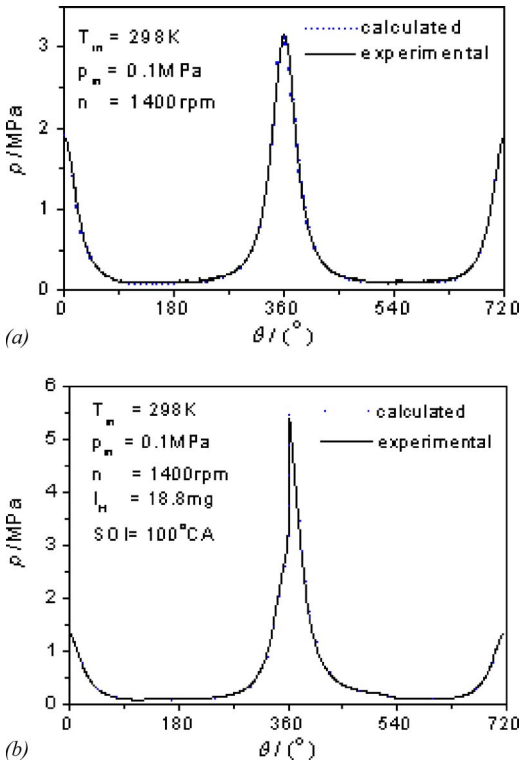


Fig. 3 Comparison of cylinder pressure: (*a*) motoring case, (*b*) HCCI case

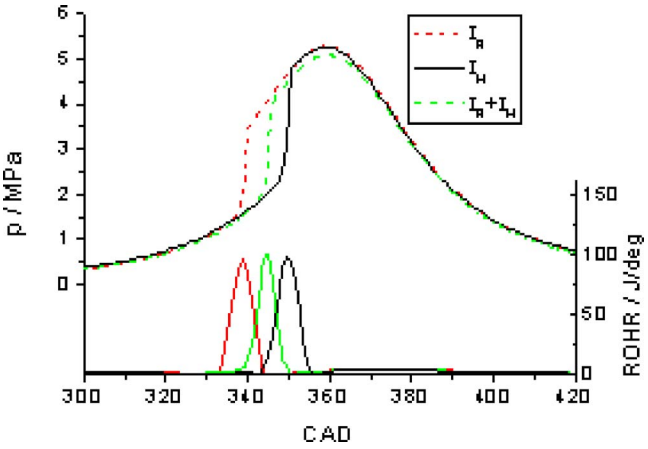


Fig. 4 Calculated cylinder pressure and heat release rate of HCCI combustion using three injection strategies (T_{in} =300 K, p_{in}=0.1 MPa, n=1400 rpm, g_b=15.5 mg)

TSDI Fuel Injection Strategies. Figure 4 shows the calculated pressure and heat release rate curves of the three different injection strategies (I_R,I_H,I_R+I_H) with same total fuel mass per cycle (g_b=15.5 mg) at HCCI(3) CAM phase. The split ratio (g_r) of the two-stage injection was 0.2. g_r is defined as the first injection mass/total fuel mass per cycle. SOI_1 and SOI_2 are start of first and second injection, respectively. The corresponding mass of injection is g_{b1} and g_{b2},

$$g_r = \frac{g_{b1}}{g_b}$$

The corresponding test conditions and engine performance of each case are listed in Table 4.

From Fig. 4, it can be seen that injection during NVO makes ignition advance significantly and combustion rate increase. When split injection is applied, ignition timing can be adjusted in a wide range between 336°CA and 347°CA. The reason for advanced ignition timing is mainly due to active species since the fuel injected during NVO can be partly oxidized and reformed as shown in Fig. 5.

In order to investigate the effects of active additives on HCCI combustion, single-zone simulations were carried out using SENKIN. Figure 6 shows the effects of different additives on HCCI combustion characteristics. It can be seen that even a small amount of active additives present in the intake charge will be important in HCCI combustion. When introducing 1000 ppm NO, H_2O_2, and CH_2O, respectively, ignition timing advances obviously. The phenomena can be used to explain why fuel reforming during NVO can improve ignition in HCCI engines with negative

Table 4 Calculated combustion characteristics

Case	I_R	I_R+I_H	I_H
Lambda	1.13	1.09	1.15
g_b (mg)	15.5	15.5	15.5
g_r	1	0.2	0
SOI_1 (°CA)	−50	−50	
SOI_2 (°CA)		100	75
MFB50% (°CA)	340	345	351
EGR ratio	42%	44%	43%
η_v	51%	50%	52%
Heatloss (kJ)	−0.224	−0.215	−0.197
Blow-by (g)	−0.0110	−0.0106	−0.0106
IMEP (MPa)	0.30	0.32	0.34
η_i	36%	39%	41%
ISFC g/(kW h)	229	214	202

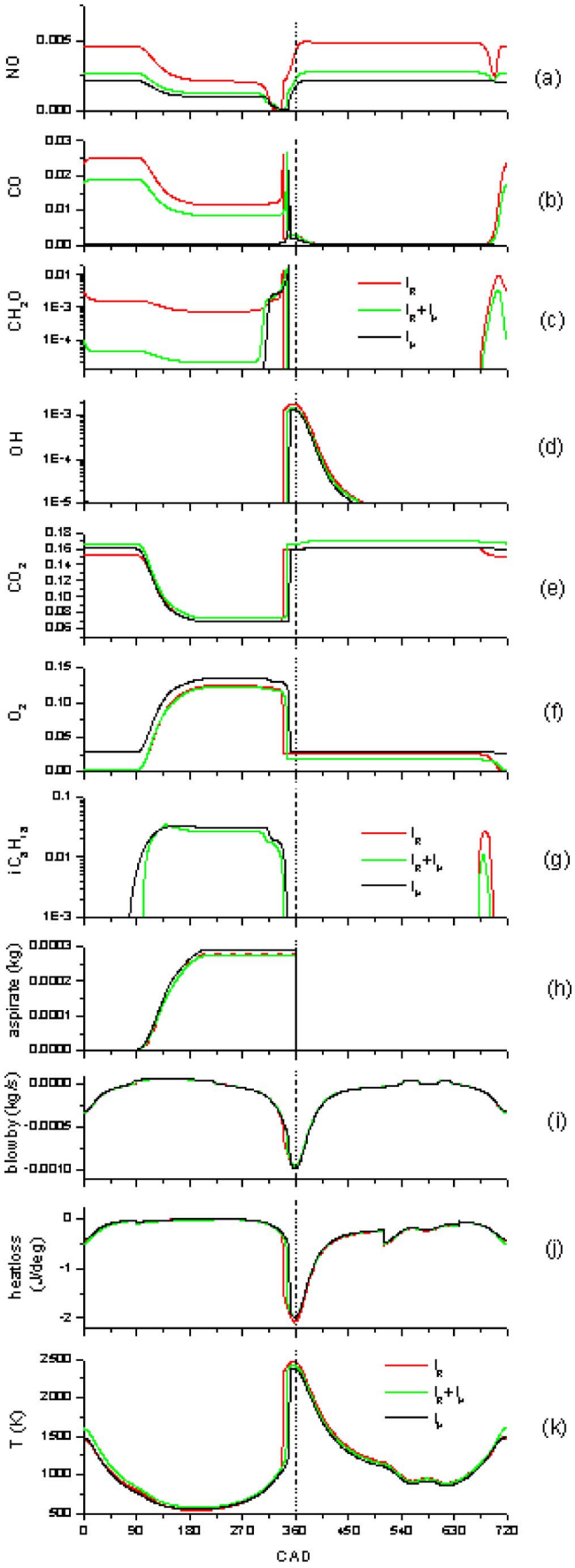


Fig. 5 Calculated main species and working process of the HCCI engine using three injection strategies

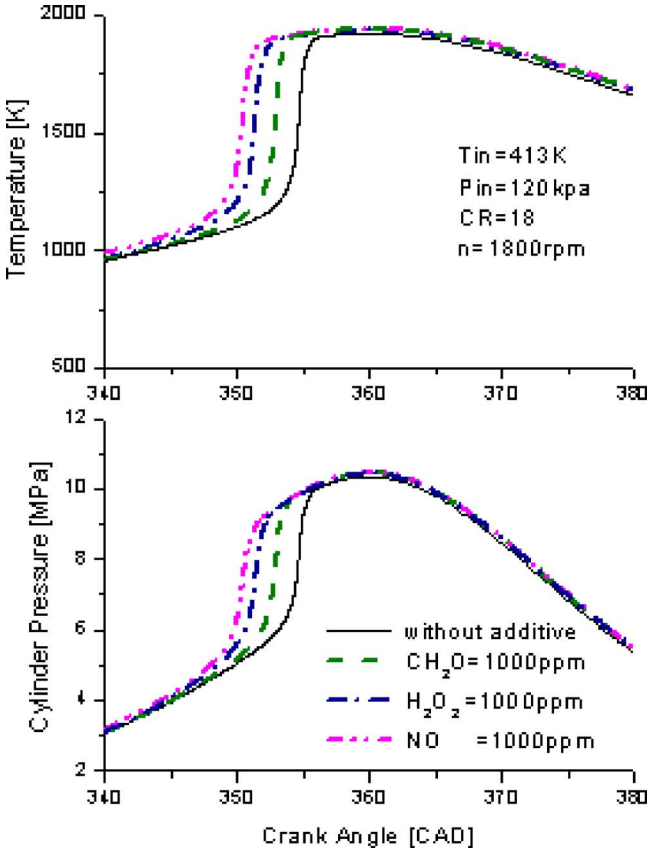


Fig. 6 Calculated cylinder temperature and pressure curves with different additives

valve overlap.

From Fig. 5, when fuel is injected during NVO, fuel is reformed and converted to HC, CO, and CH_2O. More fuel injected during NVO leads to more active species, such as CH_2O, as shown in Fig. 5(*c*). CH_2O presents immediately after injection and resides in the cylinder until ignition occurs. From the above simulated results of additive influence, introducing 1000 ppm CH_2O improves ignition significantly. Therefore, NVO injection advances ignition. From Table 4, earlier ignition before TDC leads to lower thermal efficiency combined with higher heat-loss and blow-by as shown in Figs. 5(*i*) and 5(*j*). Meanwhile, it can be seen that internal EGR ratio and volumetric efficiency (η_V) are various with injection strategy even at fixed cam phase configuration. Since fuel which decomposes during NVO leads to higher cylinder pressure at IVO, NVO injection (I_R) usually makes the intake charge decrease.

Experimental Results and Discussion

Figure 7 shows the measured pressure and heat release rate curves using three different injection strategies (I_R,I_H,I_R+I_H) with the same total fuel mass per cycle in the test engine. The I_R injection timing was −15°CA and the I_H injection timing was 100°CA. The split ratio was 0.2.

From Fig. 7, the ignition timing was 348°CA for the I_R case, and 360°CA for the I_H case. Injection during NVO makes ignition timing advance and combustion rate increase. Since fuel injected during NVO can be partly oxidized and reformed, when split injection is applied, ignition timing can be adjusted in a wide range flexibly.

The engine cycle simulation results in Fig. 5 can also prove this. The preinjection during NVO generates active species such

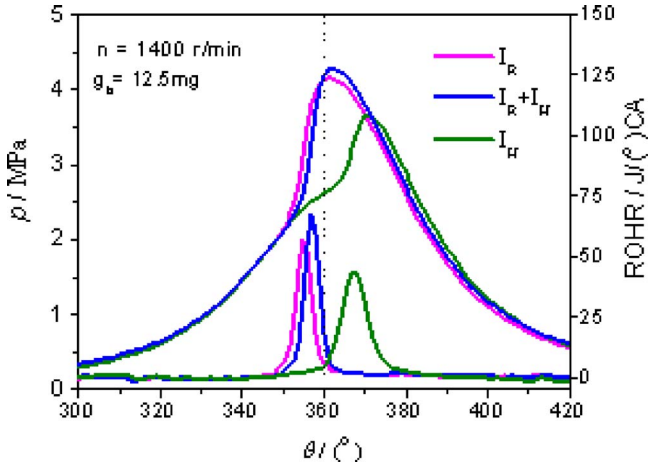


Fig. 7 Effects of injection strategies on ignition

as CH_2O, which advance ignition. When split injection is applied, the ignition timing can be controlled in a wide range between 348°CA and 360°CA.

Figure 8 shows the effects of the different split injection ratios with the same total fuel mass per cycle, g_b=15 mg, SOI_1=−45°CA, SOI_2=100°CA. With an increase of the injected fuel mass during NVO, ignition timing advances, and combustion rate increase. The reason is that fuel reforming is enhanced with more fuel injected during NVO, which advances ignition. The reform has obvious effects: Active species amount is increased, which advances heat release. Therefore, the split injection ratio can be used to control HCCI ignition timing and combustion rate.

The effects of injection timing (SOI_1) during NVO on pressure curve and heat release rate of HCCI combustion are shown in Fig. 9. g_b=14.2 mg, g_r=0.18, SOI_2=100°CA. It is indicated that the ignition timing advances with earlier injection. If the I_R injection is too early, combustion ends before TDC, which leads to rapid pressure rise with noises, more heat transfer loss, and gas blow-by. If the I_R injection is too late (near the intake TDC), knocking and misfire occurs alternately, which leads to high COV and deteriorates combustion. Therefore the preinjection should be adjusted to a suitable timing before TDC to make the engine operate with the best performance.

The effects of the I_R injection timing on emissions, COV, and fuel consumption are shown in Fig. 6. If the preinjection is too early (SOI_1=−57°CA), fuel is injected into high temperature EGR, which makes intense oxidation reactions occur before intake TDC with further increase in temperature and pressure by

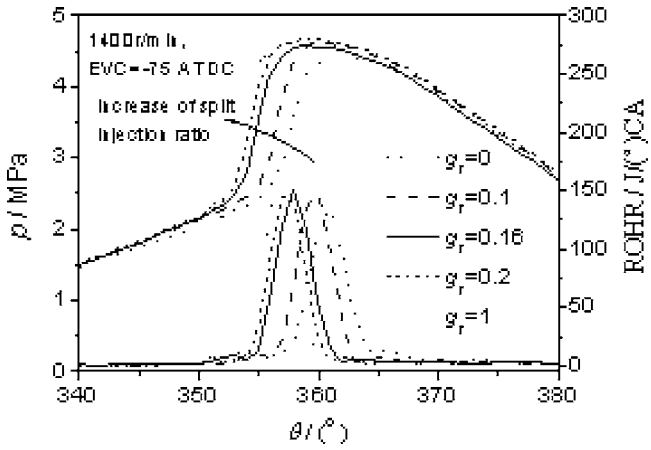


Fig. 8 Effects of split injection ratio on combustion

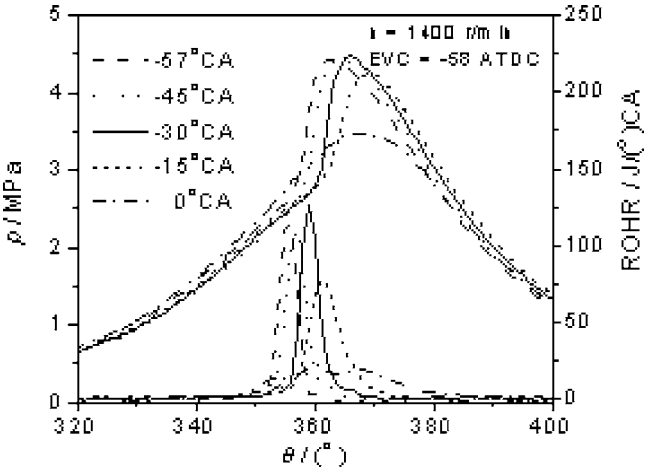


Fig. 9 Effect of preinjection timing during NVO on combustion charateristics

compression. This leads to high heat transfer and fuel consumption. Ignition advances, heat release process ends before TDC, and peak temperature in cylinder increases, which leads to higher NO_x emission. When the injection is too late ($SOI_1=0°CA$) it leads to fuel film because fuel spray impinges on the piston bowl, which increases HC and CO emissions. Misfire and knocking occurs alternately, NO_x, COV, and fuel consumption increases and engine performance is extremely deteriorated. From Fig. 10, the optimal injection time is from −45°CA to −15°CA before intake TDC.

Figure 11 shows the pressure curves of six engine operation conditions at different loads with different split injection ratios in two stages. To avoid knocking, with the increase in load, preinjection during NVO must be decreased while the main injection must be increased. Considering the practicability of the engine, the preinjection quantity during NVO was fixed at $g_{b1}=5$ mg at low load (IMEP<0.3 MPa). The engine load was increased by increasing the main-injection amount.

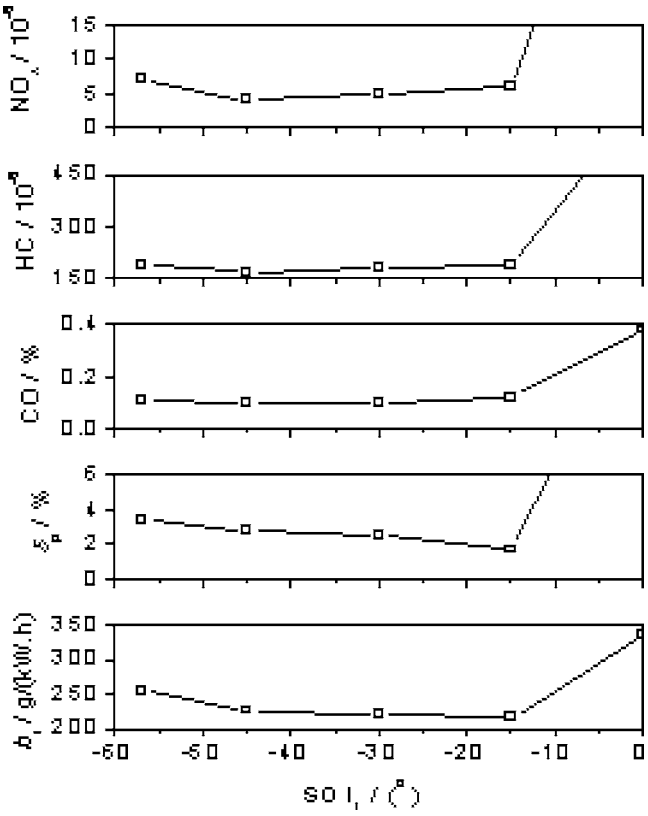


Fig. 10 Influence of preinjection timing on emissions

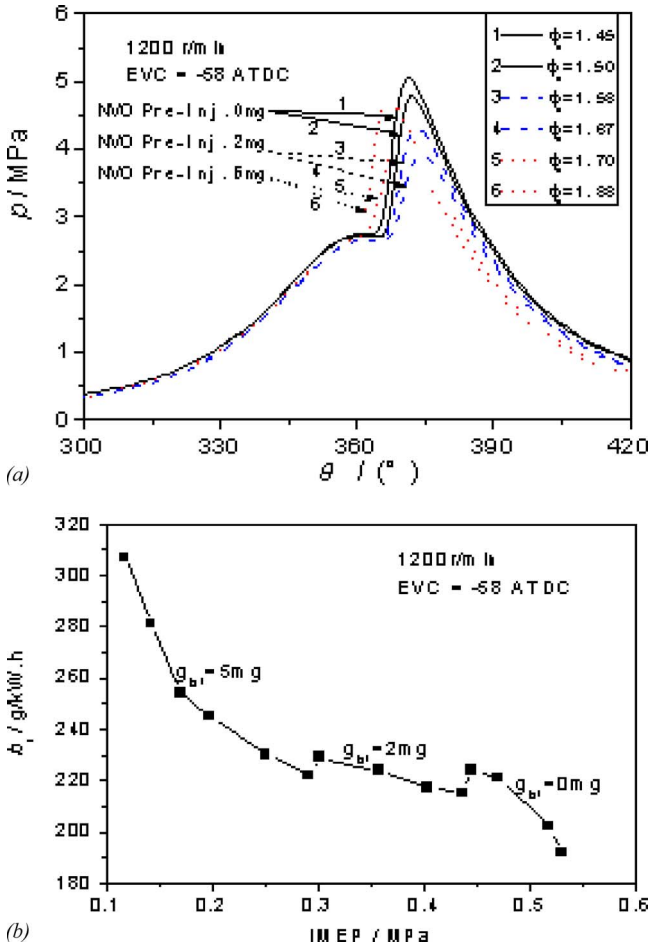


Fig. 11 Injection strategies adjusted with load; (*a*) pressure curve, (*b*) load versus fuel consumption

From Fig. 11(*a*), when light knocking occurs, decreasing preinjection mass to $g_{b1}=2$ mg can extend IMEP to 0.43 MPa. When g_{b1} during NVO decreases to zero, IMEP can reach 0.53 MPa. The indicated specific fuel consumption (ISFC) at this point is ISFC=192 g/kW h, air/fuel equivalence ratio (ϕ_a, or λ) is 1.45, and the measured NO_x emission is 89 ppm.

From Fig. 11(*b*), with the increase of load at every fixed preinjection mass, the ignition timing advances (Fig. 11(*a*)). This can be explained by the fact that higher loads result in higher temperatures of internal EGR.

Conclusions

In this paper, HCCI ignition and combustion rate was controlled flexibly by fuel reforming on a HCCI engine. The effects of injection strategies on fuel economy, emission, and heat release were studied on a four-stroke HCCI engine. It was found that:

(1) The injected fuel was reformed with the residual gas in the cylinder during NVO due to the high temperature resulting from exhaust gas recompression.
(2) Using the NVO-TSDI injection strategy, the fuel concentration and temperature as well as chemical species of the mixture in the cylinder can be controlled flexibly by adjust-

ing the split injection ratio. The ignition timing and combustion rate can be controlled directly and effectively by the split injection ratio.

(3) Fuel injection during NVO can realize fuel reformation, enhance ignition, and reduce COV of the combustion. NVO injection can be used to extend the HCCI low load limit. With the HCCI load increase, the split injection ratio should decrease to avoid knocking.

(4) The studies on the injection ratio and injection timing show that, either earlier or more preinjection leads to earlier ignition and more rapid combustion. Therefore, both the preinjection timing and split-injection ratio can be used to control gasoline HCCI combustion effectively.

Acknowledgment

This research work is supported by the National Key Basic Research Plan ("973" Plan), "Basic research on combustion theory of new generation IC engines and alternative fuels: Controlled homogeneous compression ignition of high octane fuels," under Grant No. 2001CB209204.

Nomenclature

ATDC = after top dead center
BDC = bottom dead center
BTDC = before top dead center
CAD = crank angle degree
CI = compression ignition
COV = coefficient of variation
CR = compression ratio
DI = direct injection
DOC = duration of combustion
DOI = duration of injection
$dp/d\phi$ = rate of pressure rise (MPa/deg)
EGR = exhaust gas recirculation
EVC = exhaust valve close
EVO = exhaust valve open
g_b = total fuel mass per cycle
g_r = ratio of split injection
GDI = gasoline direction injection
HCCI = homogeneous charge compression ignition
HCSI = homogeneous charge compression ignition
HRG = 0D model of high RON gasoline
IMEP = indicated mean effective pressure
ISFC = indicated specified fuel consumption
IVC = intake valve close
IVO = intake valve open
n = engine speed (r/min)
NO_x = nitrogen oxides
NVO = negative valve overlap
OH = hydroxyl
PFI = port fuel injection
P_{in} = intake pressure
RCCI = radical charge compression igntion
ROHR = rate of heat release
RON = research octane number
SCCI = stratified charge compression ignition
SCSI = stratified charge spark ignition
SI = spark ignition
SOC = start of combustion
SOI = start of injection
TDC = top dead center
TSDI = two-stage direct injection
T_{in} = intake temperature
WOT = wide open throttle
η_i = thermal efficiency
η_V = volumetric efficiency
ε = compression ratio
ϕ = fuel/air equivalence ratio
ϕ_a = air/fuel equivalence ratio

References

[1] Zhao, F., Asmus, T. W., Assanis, D. N., Dec, J. E., Eng, J. A., and Najt, P. M., 2003, *Homogenous Charge Compression Ignition (HCCI) Engine: Key Research and Development Issues*, Society of Automotive Engineers, Warrendale, PA.

[2] Najt, P. M., and Foster, D. E., 1983, "Compression-Ignited Homogeneous Charge Combustion," SAE Paper No. 830264.

[3] Noguchi, M., Tanaka, Y., Tanaka, T., and Takeuchi, Y., 1979, "A Study on Gasoline Engine Combustion by Observation of Intermediate Reactive Products During Combustion," SAE Paper No. 790840.

[4] Iida, N., 1997, "Alternative Fuels and Homogeneous Charge Compression Ignition Combustion Technology," SAE Paper No. 972071.

[5] Aceves, S. M., Flowers, D. L., Westbrook, C. K., Smith, J. R., Pitz, W., Dibble, R., Christensen, M., and Johansson, B., 2000, "A Multi-Zone Model for Prediction of HCCI Combustion and Emissions," SAE Paper No. 2000-01-0327.

[6] Fiveland, S. B., and Assanis, D. N., 2002, "Development and Validation of a Quasi-Dimensional Model for HCCI Engine Performance and Emissions Studies Under Turbocharged Conditions," SAE Paper No. 2002-01-1757.

[7] Ogink, R., and Golovitchev, V., 2001, "Gasoline HCCI Modeling: Computer Program Combining Detailed Chemistry and Gas Exchange Processes," SAE Paper No. 2001-01-3614.

[8] Milovanovic, N., and Chen, R., 2004, "Influence of the Variable Valve Timing Strategy on the Control of a Homogeneous Charge Compression (HCCI) Engine," SAE Paper No. 2004-01-1899.

[9] Yang, J., Culp, T., and Kenney, T., 2002, "Development of a Gasoline Engine System Using HCCI Technology—The Concept and the Test Results," SAE Paper No. 2002-01-2832.

[10] Hyvönen, J., Haraldsson, G., and Johansson, B., 2003, "Supercharging HCCI to Extend the Operating Range in a Multi-Cylinder VCR-HCCI Engine," SAE Paper No. 2003-01-3214.

[11] Christensen, M., Hultqvist, A., and Johansson, B., 1999, "Demonstrating the Multi Fuel Capability of a Homogeneous Charge Compression Ignition Engine With Variable Compression Ratio," SAE Paper No. 1999-01-3679.

[12] Oakley, A., Zhao, H., and Ladommatos, N., 2001, "Dilution Effects on the Controlled Auto-Ignition (CAI) Combustion of Hydrocarbon and Alcohol Fuels," SAE Paper No. 2001-01-3606.

[13] Olsson, J.-O., and Erlandsson, O., 2000, "Experiments and Simulation of a Six-Cylinder Homogeneous Charge Compression Ignition (HCCI) Engine," SAE Paper No. 2000-01-2867.

[14] Eng, J. A., Leppard, W. R., and Sloane, T. M., 2003, "The Effect of Di-Tertiary Butyl Peroxide (DTBP) Addition to Gasoline on HCCI Combustion," SAE Paper No. 2003-01-3170.

[15] Kopecek, H., and Wintner, E., 2004, "Laser-Stimulated Ignition in a Homogeneous Charge Compression Ignition Engine," SAE Paper No. 2004-01-0937.

[16] Urushihara, T., Hiraya, K., Kakuhou, A., and Itoh, T., 2003, "Expansion of HCCI Operating Region by the Combination of Direct Fuel Injection, Negative Valve Overlap and Internal Fuel Reformation," SAE Paper No. 2003-01-0749.

[17] Sun, Y., Shuai, S.-J., Wang, J.-X., Wang, Y.-J., 2003, "Numerical Simulation of Mixture Formation and Combustion of Gasoline Engines with Multi-Stage Direct Injection Compression Ignition (DICI)," SAE Paper No. 2003-01-1091.

[18] Sjöberg, M., Edling, L.-O., Eliassen, T., Magnusson, L., and Angström, H.-E., 2002, "GDI HCCI: Effects of Injection Timing and Air Swirl on Fuel Stratification, Combustion and Emissions Formation," SAE Paper No. 2002-01-0106.

[19] Wang, Z., Wang, J.-X., and Shuai, S.-J., 2005, "Effects of Spark Ignition and Stratified Charge on Gasoline HCCI Combustion with Direct Injection," SAE Paper No. 2005-01-1037.

[20] Wang, Z., Wang, J.-X., and Shuai, S.-J., 2004, "Numerical Simulation of HCCI Engine With Multi-stage Gasoline Direct Injection Using 3D-CFD With Detailed Chemistry," SAE Paper No. 2004-01-0563.

[21] Chiodi, M., and Bargende, M., 2001, "Improvement of Engine Heat-Transfer Calculation in the Three-Dimensional Simulation Using a Phenomenological Heat-Transfer Model," SAE Paper No. 2001-01-3601.

[22] Koopmans, L., Ogink, R., and Denbratt, I., 2003, "Direct Gasoline Injection in the Negative Valve Overlap of a Homogeneous Charge Compresson Ignition Engine," SAE Paper No. 2003-01-1854.

An Improved Soot Formation Model for 3D Diesel Engine Simulations

Joan Boulanger[1]
e-mail: joan.boulanger@nrc-cnrc.gc.ca

Fengshan Liu

W. Stuart Neill

Gregory J. Smallwood

Institute for Chemical Process and Environmental Technology,
1200 Montréal Building M-9,
Ottawa, ON, K1A0R6, Canada

Soot formation phenomenon is far from being fully understood today and models available for simulation of soot in practical combustion devices remain of relatively limited success, despite significant progresses made over the last decade. The extremely high demand of computing time of detailed soot models make them unrealistic for simulation of multidimensional, transient, and turbulent diesel engine combustion. Hence, most of the investigations conducted in real configuration such as multidimensional diesel engines simulation utilize coarse modeling, the advantages of which are an easy implementation and low computational cost. In this study, a phenomenological three-equation soot model was developed for modeling soot formation in diesel engine combustion based on considerations of acceptable computational demand and a qualitative description of the main features of the physics of soot formation. The model was developed based on that of Tesner et al. and was implemented into the commercial STAR-CD™ CFD package. Application of this model was demonstrated in the modeling of soot formation in a single-cylinder research version of Caterpillar 3400 series diesel engine with exhaust gas recirculation (EGR). Numerical results show that the new soot formulation overcomes most of the drawbacks in the existing soot models dedicated to this kind of engineering task and demonstrates a robust and consistent behavior with experimental observation. Compared to the existing soot models for engine combustion modeling, some distinct features of the new soot model include: no soot is formed at low temperature, minimal model parameter adjustment for application to different fuels, and there is no need to prescribe the soot particle size. At the end of expansion, soot is predicted to exist in two separate regions in the cylinder: in the near wall region and in the center part of the cylinder. The existence of soot in the near wall region is a result of reduced soot oxidation rate through heat loss. They are the source of the biggest primary particles released at the end of the combustion process. The center part of the cylinder is populated by smaller soot particles, which are created since the early stages of the combustion process but also subject to intense oxidation. The qualitative effect of EGR is to increase the size of soot particles as well as their number density. This is linked to the lower in-cylinder temperature and a reduced amount of air. [DOI: 10.1115/1.2718234]

Keywords: soot modeling, diesel engine, emission, particle, automotive engineering

Introduction

The merits of diesel engines, compared to other internal combustion engines, are lower fuel consumption, and unburned hydrocarbons, due to the overall lean combustion (equivalence ratio of the order of 0.5), and a better energy release efficiency due to a controlled nonhomogeneous combustion (diffusion flame) at high pressures. Diesel engine is therefore an attractive option to reduce CO_2 emissions from automobiles and counter greenhouse gas effects [1]. On the other hand, because of the existence of rich high temperature zones leading to fuel pyrolysis, a diesel engine is very likely to produce particulates emission in the exhaust gas. Given the intensive use of diesel engines and the detrimental effects of soot particulates on environment and health [2], more stringent emissions standards have been imposed [3], which challenge the viability of diesel engines.

Soot formation phenomenon is far from being fully understood despite the significant progress in fundamental understanding made in the last two decades. Simple engineering correlations and simplified soot formation models available for simulation of soot in practical combustion devices remain of relatively limited success. In the early 1970s, Khan and Greeves [4] presented the first model for the soot production from diesel engines. The most complete models describing soot dynamics formation use chemical kinetics-like approach, as in Refs. [5–8]. However, because of the numerous Arrhenius terms appearing in these models, it is questionable to use this type of modeling for simulation of multidimensional, transient, and turbulent diesel engine combustion. Beside this "kinetic approach," "empirical approaches" are also widely used, particularly for soot formation prediction in industrial configurations. The majority of the phenomenological approaches belong to the one-step fuel based models [4,9,10]. Indisputably, most of the modern numerical studies on multidimensional diesel engine computation with prediction of soot emission [1,11–20] have been made with the help of those empirical models, which consist of only one Arrhenius term in the soot formation step with two empirical constants: a moderately high activation temperature and a preexponential constant. The advantages of these empirical soot models are easy implementation and low computational costs but with the drawbacks of an often poor representation of the physical and chemical processes.

[1]Present address: Gas Turbine Laboratory–Institute for Aerospace Research, National Research Council of Canada, Building M-10 Room 104, 1200 Montréal Road, K1A 0R6 Ottawa, ON, Canada.

Contributed by the Internal Combustion Engine Division of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received January 23, 2006; final manuscript received December 13, 2006. Review conducted by Jim Cowart. Paper presented at the 2005 Fall Conference of the ASME Internal Combustion Engine Division (ICEF2005), September 11–14, 2005, Ottawa, ON, Canada.

These models thus in general suffer a narrow range of applicability around the conditions under which they were developed.

It is well established that soot formation is divided into two major steps: *inception* and *growth*, both following the pyrolysis of fuel, a decomposition and incomplete combustion of hydrocarbons (HC). Each of them may be further divided into two substeps. First, inception is related to the formation of soot precursors, whose nature is not clearly defined yet, but generally believed to be sufficiently large polycyclic aromatic hydrocarbons (PAHs) [7], produced as intermediate products directly from fuel pyrolysis and then conversion through polymerization into particulate nuclei when a sufficient mass is reached. This is also the initial stage that a physical surface appears. Second, soot particulate surface growth proceeds through the addition of carbon atoms by heterogeneous reaction at the surface of the particulate with the pyrolysis products and coagulation. Eventually, when they get older, primary particles assemble in fractal clusters. Clearly, actual soot formation models dedicated to industrial configuration simulations are unable to capture all aspects of the phenomena. However, given the new emission standards, it becomes important to better characterize the soot formation process in order to improve the design of an engine by incorporating efficiently the key knowledge on soot dynamics to meet the current and forthcoming regulations on particulate emission.

It is intended in this paper to develop a computationally efficient phenomenological soot model within the context of diesel engine combustion. The following two criteria were sought during the development of the present model: (i) capture of the main physical processes of soot formation, and (ii) without significant increase in computing time or the complexity of the formalism. The soot model developed in this study was incorporated into the STAR-CD™ package. Its capability was demonstrated in particulate emission prediction from a single-cylinder research version of the Caterpillar™ 3400 series diesel engine with exhaust gas recirculation (EGR).

Development of a Three-Equation Model

The present model is derived from the Tesner et al. model [21]. This model has the advantage of embedding the representative generic steps related to soot particle formation in a system of two Semenov equations. The interest is in complementing this model by a soot particle surface growth process, which can help establishing an explicit link between fuel concentration and soot mass growth. This alleviates the requirement for an ad hoc prescribed diameter of the primary soot particles in the original model of Tesner et al. (given a soot particle number density distribution in the cylinder volume, prescription of the size results in a major nuisance of the soot mass distribution without any consideration for the physics of soot formation/combustion dynamics). The other expected benefits are a reduced tuning process—for most of the parameters may be identified with the help of fundamental data—and a good computation efficiency since the equations are kept relatively simple and not too nonlinear, compared to the existing models. It is thus hoped that most of the interesting features of soot formation dynamics can be predicted at a low computational cost, convenient for engineering applications. Given the relatively poor knowledge on cluster formation, the model is only able to predict the primary particle size and history but not soot aggregates.

Regulatory agencies are more and more interested in knowing the emitted particle size from industrial furnaces or engines [2,3], rather than the soot mass only, for the following two reasons. First, the smaller particles are actually more deleterious to human health and environment [22,23]. Second, although modern diesel engines have lower particulate emissions, they often emit particles that are smaller in size but larger in number density, as a result of the technical solutions adopted to decrease the overall soot mass production. Emission control systems for particulate matter are not always efficient in retaining all soot particles, depending on their respective size (<50 nm) [24]. Diesel particle filters rely on two distinct technologies: *strain filtration* or *deep bed filtration*. In case of strain filtration, the consequences of small particles are obvious as they escape through the perforated medium acting as a filter and are freely released to the atmosphere. However, even deep bed filtration technology efficiency is damaged for small particles have less chance of collision with the filter when they go through the SiC fibers. Attempts of predicting the primary particle size characteristic of the emission represent a challenge for numerical predictive tools dedicated to diesel engine design and are the first step before considering the cluster formation modeling, which form the effective entities to be trapped.

Several models simulate the surface growth rate of soot particles through interactions between the surface density of soot primary particles and the surrounding fuel vapor [5,25] (or its break-down products [6,8]). The usual form of this growth process may be written as

$$K_G N[\text{Fuel}]A_s \tag{1}$$

which describes the heterogeneous reaction of the surrounding fuel molecules at the surface A_s of the soot primary particles. Here NA_s is the volumetric density of available soot particle surface area, associated with a locally monodispersed primary particle distribution. In Eq. (1), [Fuel] is the molar concentration of the hydrocarbon fuel and K_G is physically observed as a time-decaying constant modeling the saturation and stabilization of the active sites on soot particle surface during its growth. Such a time function is not convenient to implement and the models used in Refs. [5,6,8,25,26] make use of the undesirable Arrhenius term. Another somewhat crude way of handling the aging effect of soot particles is to consider it as inversely proportional to the area of soot particles, i.e., $K_G = K_{G,\text{incep}}/A_s$ ($K_{G,\text{incep}}$ represents the constant associated with inception soot particles—it is different from K_G as it includes A_s), thus canceling the surface dependence of the surface growth rate given in Eq. (1). In Ref. [26], the surface dependence was reduced to its square-root for a similar reason. In the present study, the assumption that K_G is inversely proportional to A_s is made in favor of simplicity, as the goal is to use a minimum set of rate equations that can describe the most important soot phenomena. Constant $K_{G,\text{incep}}$ is derived from the kinetic theory describing the collision frequency between Brownian particles (here the original hydrocarbon molecules) and the inception primary soot particles, assumed of spherical shape with a size and an inertia much larger than the hydrocarbon molecules. It was also assumed that the collision between a hydrocarbon molecule and an inception primary soot particle yields the release of all carbon atoms in the original fuel hydrocarbon molecule [5] and this assumption is also made in the present formulation. Hence, $K_{G,\text{incep}}$ is written as

$$K_{G,\text{incep}} = \sqrt{\frac{8RT}{\pi \bar{M}_F}} \frac{\pi D_0^2}{4} \bar{M}_c m \tag{2}$$

where $\bar{M}_F$ and $\bar{M}_c$ are the molar weights of fuel and carbon, respectively, D_0 is the inception diameter of soot particles. m represents the number of carbon atoms in the hydrocarbon molecule.

As mentioned earlier, soot particle growth is due to heterogeneous reaction involving growth species which are obtained from fuel breakdown. Directly linking soot growth to fuel concentration is obviously an approximation. Given that, to first order, fuel and HC-growth species coexist in space, it is assumed that soot growth is satisfactorily mimicked by reactions with primary fuel molecules, as has already been done for Fenimore reburn NO mechanism [27].

Furthermore, inception of soot particle from radicals also contributes to an initial amount of carbon matter and can be evaluated by $C_a a n$ with $C_a = \rho_s (\pi D_0^3)/6$. The equation for particle formation in Ref. [21] can thus be split into the following two equations:

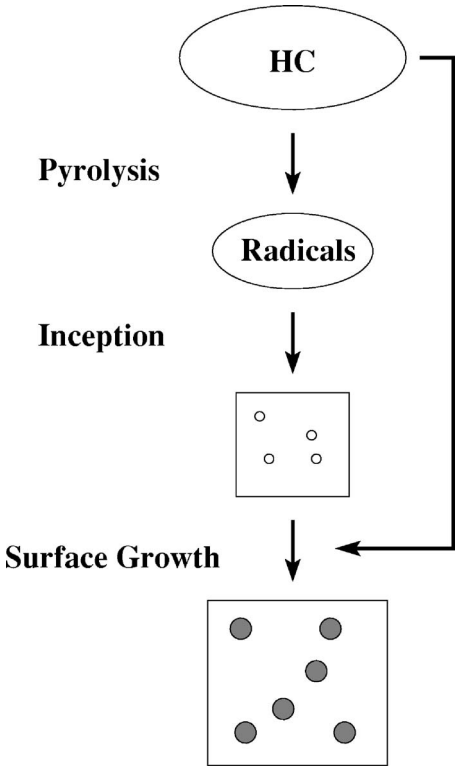


Fig. 1 Flow chart of the baseline soot model. HC is considered as the active specie at each step of the soot formation process for sake of simplicity and efficiency.

$$\frac{dN}{dt} = an - bNn \tag{3}$$

$$\frac{d(\rho y_s)}{dt} = K_{G,\text{incep}} N[\text{Fuel}] + C_a an - NA_s S_{\text{ox}} \tag{4}$$

that respectively describe the production rate of soot particle number density and soot mass. After taking into account the coagulation that reduces the soot particle number density, as suggested in Ref. [6], Eq. (3) takes the following form:

$$\frac{dN}{dt} = an - bNn - K_c\sqrt{T}\left(\frac{\rho y_s}{\rho_s}\right)^{1/6} N^{11/6} \tag{5}$$

with K_c being the coagulation coefficient.

The radical formation equation from Ref. [21] is retained here:

$$\frac{dn}{dt} = a_0 N_F \exp\left(-\frac{T_{a_{n_o}}}{T}\right) + Fn - g_0 Nn - S'_{\text{ox}} \tag{6}$$

Equations (4)–(6) constitute the three-equation soot model formulated in the present study. Figure 1 illustrates the physical and chemical processes leading to soot particles assumed in the formulation of the present soot model. First, as exhibited through the first term of Eq. (6), pyrolysis leads to the formation of radicals from fuel molecule cleavage. Those unstable radicals may increase in number through chain branching (second term) and may be destroyed when landing on soot particles (the last term). This is the original Semenov equation developed by Tesner et al. for the radicals [21]. As exhibited by the first term of Eq. (5) and the second term of Eq. (4), the presence of radicals leads to the assembling of nascent solid particles (diameter about 1 nm). Today's school of thinking interprets this with the help of PAH polymerization process. Once soot particles are formed, their growth is controlled by surface reaction and coalescence, the first and last terms of Eqs. (4) and (5), respectively. In the following, the reader will find a brief description of each term and the value assigned.

S_{ox} in Eq. (4) is the sum of the contribution of the following soot oxidation models. The oxidation by O_2 is given by the Nagle-Strickland-Constable [28] model (S_{NSC}). Soot oxidation by OH and O attacks are taken into account according to the Fenimore and Jones mechanism [29] (S_{FJ}) and the relation found in Ref. [30] (S_{BDEM}), respectively. Thus, we have

$$S_{\text{ox}} = S_{\text{NSC}} + S_{\text{FJ}} + S_{\text{BDEM}} \tag{7}$$

with

$$S_{\text{NSC}} = 1.2 \times 10^2 \times \left\{\frac{k_A P_{O_2}\chi}{1 + k_z P_{O_2}} + k_B P_{O_2}(1-\chi)\right\} \tag{8}$$

$$S_{\text{FJ}} = 1.27 \times 10^3 \Gamma_{\text{OH}} P_{\text{OH}} T^{-1/2} \tag{9}$$

$$S_{\text{BDEM}} = 665.5 \Gamma_{\text{O}} P_{\text{O}} T^{-1/2} \tag{10}$$

Symbols in Eqs (8)–(10) are specific constants of the models whose clear definition is detailed in the respective references. P_i is the partial pressure of specie "i." The collision efficiencies Γ_i are: $\Gamma_{\text{O}}=0.5$ and $\Gamma_{\text{OH}}=0.42\tanh(2664/T-2.8)+1)$.

Oxidation is included in the model with an assumption that it does not directly affect the soot particle number density N. However, a soot primary particle is considered as destroyed when its diameter falls below the inception diameter D_o, mimicking implosion of the smallest primary particles due to the volume oxidation [31]. The concentrations of O and OH radicals are estimated based on partial-equilibrium relations [32].

Given the poor knowledge on the nature of the generic radicals, the oxidation of the latter is treated using the Magnussen and Hjertager model [33] modified as in [17], and is based only on O_2,

$$S'_{\text{ox}} = An\frac{\varepsilon}{\kappa}\min\left(1, \frac{y_{O_2}}{y_s \upsilon_s + y_F \upsilon_F}\right)(1 - e^{-T/1800}) \tag{11}$$

A is the constant of the model. κ/ε is the characteristic turbulence time. y_{O_2} and y_F are the di-oxygen and fuel mass fraction, respectively. ν_s and ν_F are the stoichiometric oxygen requirements to burn 1 kg of soot and fuel, respectively. Radicals, soot, and fuel are present in this oxidation rate for radicals are they are all in competition to access oxygen.

The parameters of the present soot model are summarized in the following (in SI units):

- $T_{a_{n_o}} = 21{,}000$ K. This activation temperature is correlated to the energy bond break [21] and is consistent with the oxidative pyrolysis [31]. It is expected to be fuel-dependent, in relation with the degree of saturation of the hydrocarbon molecule. The discussion in Ref. [26] for hydrocarbon mixture (p. 291) has been followed here.
- $a=10^5$ was suggested in Ref. [21], based on the characteristic time of primary particles assembling from radicals.
- $F=100$, $g_0=10^{-15}$, $b=8\times10^{-14}$ were taken from Tesner et al. [21] as a result of experimental-based inverse analysis for a radical chemistry modeled using Semenov equations. It should be noted that F vanishes when the radical production rate becomes small.
- Constant a_0 is usually related to the vibration frequency of the bond to be broken. Quantum mechanics builds a bridge between this frequency and the peak in the IR spectrum for simple molecules. Obviously, this theoretical approach is subject to large uncertainty for larger hydrocarbon molecules. Due to the lack of better information, a_0 is arbitrarily set at 2.3×10^{-3} Hz in this study. It should be mentioned that a_0 is the *only* parameter which needs to be arbitrarily adjusted. It has been actually found that prediction trends

are relatively smooth with respect to a_o, helpful to a fast tuning process.
- $\rho_s = 1900$ kg/m^3 is retained for soot density.
- $D_0 = 1$ nm is the diameter of the inception soot particle.

Soot and radicals are considered here as trace species and do not influence flow properties. The turbulent Schmidt number for soot transport is assumed to be 15, since the primary particles inertia is much larger than gaseous molecules. Turbulent transport is assumed to dominate all other modes.

Numerical Methods

The three-equation soot model was implemented into the STAR-CD™ CFD code. Since this CFD package has been widely used in engine modeling community, only a brief summary of some features is given in the following.

STAR-CD™ solves the compressible, turbulent, three-dimensional transient conservation equations for reacting multi-component gas mixtures with the flow dynamics of an evaporating liquid spray treated as Lagrangian statistical polydispersed parcels on finite-volume grids. Throughout this study, the original model's constants documented in Ref. [34] were used, unless otherwise stated.

The injection process includes the modeling of the flow in the nozzle hole and atomization. The atomization model was that given by Huh [34,35]. The injector pressure is virtually set at 190 MPa. However, the velocity profile versus time is the effective quantity specified in the simulation. It is extrapolated, given the requirement of the prescribed fuel consumption, from the experimental curve provided in Ref. [17] where it is seen that the profile has a nominal exit velocity of 580 m/s. This latter value yields a nozzle discharge coefficient of about 0.83, which is consistent with the most up-to-date real devices.

The spray is modeled by using the Reitz and Diwakar model [36] and has demonstrated strong consistency with experiments on jet penetration at diesel-like conditions [37]. Perfect rebound is assumed when impinging the cylinder walls [1].

The fuel is approximately represented by dodecane for thermodynamic properties (density, viscosity, etc.) and by *n*-heptane for the chemical properties, since diesel combustion models are usually calibrated for that specie. The built-in data are used [34]. Nevertheless, the saturation pressure is enhanced in order to reach the correct range of peak temperature and ignition delay, which is justified in light of the complex nature of real diesel fuels, i.e., lighter components evaporate earlier and participate to the first stages of combustion [14].

The initial mixture at the beginning of the simulation (Intake Valve Closure, 147° BTDC) corresponds to ambient air at intake conditions ($T = 307$ K, $P = 235$ kPa). It should be noted that the residual gas in the cylinder was not taken into account in the present simulation.

The compressible RNG turbulence model [38,34] is used as it has been shown that this model is well suited to account for turbulence dynamics in an engine cylinder [12,14]. Standard wall functions [34,39,40] are applied to momentum and heat transfer at wall boundaries. No crevice flow model is used.

The Shell multistep kinetics model is applied for auto-ignition. A "laminar-turbulent" characteristic time combustion model is activated to model the turbulent combustion after ignition.

Results and Discussions

The engine numerically simulated in this study is a single-cylinder version of Caterpillar™ 3400-series (3406) heavy-duty diesel engine. The engine has four valves, a displacement of 2.44 L, electronically controlled fuel injection, and produces 74.6 kW at 1800 rpm. Further details of the engine configuration are summarized in Table 1. This engine has been well characterized in experimental and computational studies [11–18,41].

Table 1 Geometric and physical parameters of the Caterpillar 3400 series diesel engine used in the numerical simulation

Bore	137.2	mm
Stroke	165.1	mm
Connecting rod length	263	mm
Displacement	2.44	L/cyl
compression ratio	15.1	
Piston crown	Mexican hat	
Squish height	4.221	mm
Engine speed	1740	rpm
Intake pressure	235	kPa
Intake temperature	307	K
Intake valve close timing	147	deg BTDC
Swirl ratio	Nominal	
Fuel injected	3.06×10^{-5}	kg
Injection pressure	190	MPa
Nozzle diameter	0.259	mm
Spray angle from head	27.5°	
Injector protrusion	3.2	mm

Experimental measurement of particulate emissions from this engine with and without EGR were carried out by Neill and Chippior [42]. The numerical calculations were conducted under similar conditions to the experiments.

The soot model developed in this study was validated in the calculations of soot emissions from the above-described research engine against experimental data. During the course of this numerical study, the engine was not fully equipped and crucial parameters to assess a model in detail, such as cylinder pressure, heat release, injection shape were missing. The AVL test case number seven thus appeared as the easiest one to set-up through a generic approach including: top-hat injection profile, no residual gas, and generic cylinder wall temperature. We are thus primarily interested in generating new information on pollutant formation. Details of the combustion chamber geometry are provided in Fig. 2. Due to the cyclic nature of the configuration around the cylinder axis, an unstructured mesh of 60° angle sector with a moving piston and containing a single nozzle located on the bisecting plane is simulated.

Figure 3 displays the history of the total in-cylinder soot mass (m_{total}), total soot particle number (N_{total}), and the characteristic soot primary particle diameter calculated from $\rho_s \pi \overline{D^3}/6 \times N_{\text{total}} = m_{\text{total}}$. The open circles are the experimental soot emission data from Ref. [42]. The soot emission data in the case of no EGR were used to adjust the value of a_0, which is the only parameter that is arbitrarily "tuned" and is expected to be fuel dependent. The one-equation Hiroyasu soot model [10], which has been widely used in diesel engine simulation [1,11–20], was also employed in the present calculations using the constants given in Ref. [17]. Although both the Hiroyasu and the present soot models are capable of predicting the correct trend of the effect of cooled EGR on soot emission, the present model offers insight into the effect of EGR on the soot particle number and size that otherwise cannot be gained in the Hiroyasu model. This is obtained by incorporating the inception from radicals as well as the heterogeneous reaction at the surface of the soot particle in a logical sequence with respect to physical phenomena. For instance, surface reactions clearly allow predicting an increase in the soot diameter [Fig. 3(*c*)], in case of EGR. It is noted that the existence of soot particles at TDC in this case is simply due to the re-injection of the soot-containing exhaust gas (without filter) into the intake. Surface reactions are a competition between surface growth, the first term in Eq. (4), and surface oxidation. With EGR, fuel is available for a longer period of time in the cylinder, which favors surface growth [as well as a stronger creation of particles, on Fig. 3(*b*)] and the temperature is lower, weakening surface oxidation. Furthermore, the incorporation of a step corresponding to the induction radicals involves a delay in soot formation [early time in Fig. 3(*a*)], which is consistent with experiment but is badly represented by the one-step model. This is explained by the low

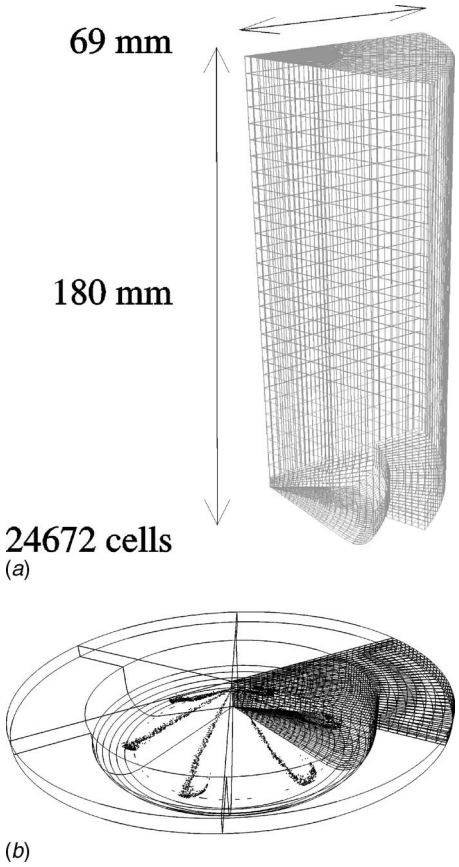


Fig. 2 CFD mesh of the cylinder geometry. Top panel: Full mesh sector (the sector picture has been stretched to show details on the layered part of the mesh subject to connectivity change and squeeze). Bottom panel: Full cylinder near TDC with spray impingement on the bowl.

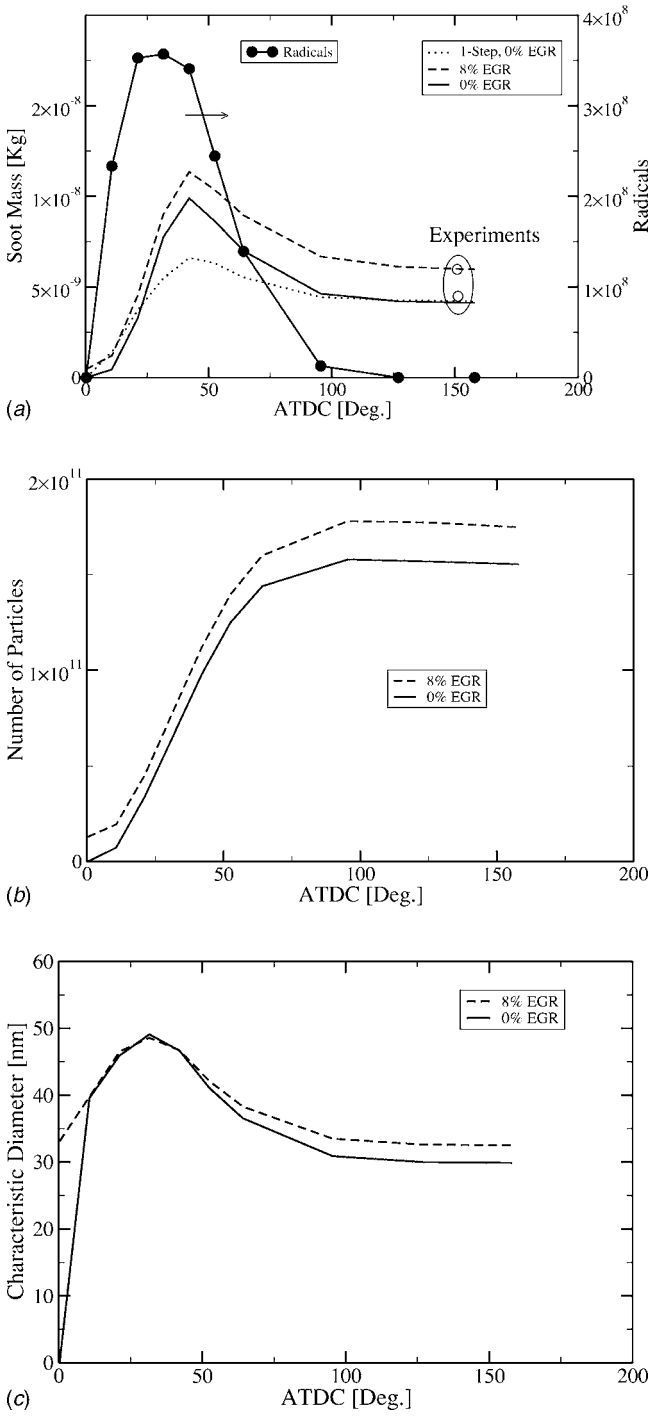


Fig. 3 History of the soot formation. (*a*) Soot mass. (*b*) Primary particle number. (*c*) Characteristic diameter. Line: present model. Dots: One-step model. Dashed line: Present model with 8% EGR. Line-circle: Induction radicals. Open circle: Experiment.

temperature dependence in the one-step model. The effect of this low temperature dependence may appear of secondary importance regarding the discrepancies induced in the overall soot production, as seen in Fig. 3(*a*). Nevertheless, the underlying consequence is the creation of soot mass in areas of the cylinder where it is physically prevented because the temperature was too low for pyrolysis/inception. Previous studies have demonstrated the importance of soot creation dynamics within the volume [8,43,44]. For instance, in Fig. 4, the soot contour in the median plan of the jet at the end of the injection is given for both models. It is seen that the one-step model predicts that soot is formed very early and close to the injector, a bias already pointed out in Ref. [45].

Coagulation is supposed to be weak, as predicted by the present model, Fig. 3(*b*). It has been suspected for two to three decades that the Brownian theory is insufficient to predict coagulation (and also agglomeration) in a highly turbulent in-cylinder flow [46,47]. However, only few very empirical approaches are currently available.

Figure 5 shows the predicted distribution of the primary particle size at the exhaust. This distribution is estimated on a cell by cell basis so that the characteristic diameter D_i of the particles contained in the cell i is given by

$$D_i = \left(\frac{6\rho y_{s_i}}{\pi \rho_s N_i} \right)^{1/3} \tag{13}$$

The weight associated with this diameter is proportional to the number of particles predicted within the cell.

Several observations are to be made. (i) The distribution of primary particle diameter lies in the acknowledged range for the diesel soot primary particle diameter, i.e., 10–50 nm. (ii) The overall diameter is increased through EGR, consistent with previous results. (iii) The distribution grossly presents two peaks. These two peaks may be explained by the existence of two soot primary particle populations inside the cylinder. These two populations are depicted in Fig. 5(*b*), where it is seen that soot with diameters below 20 nm (grossly, the left side of the distribution) are concentrated in the center of the cylinder and in the bowl while particles with a diameter larger than 40 nm are seen close to

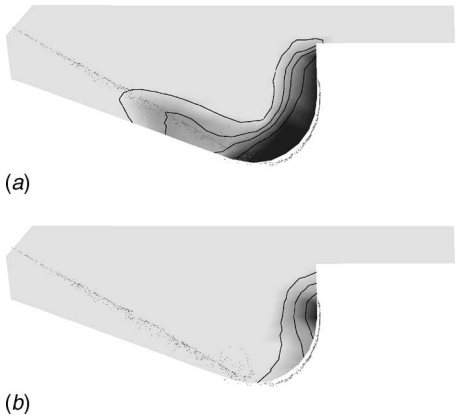


Fig. 4 Soot mass fraction distribution in the median plane of the fuel jet at the end of the injection. (*a*) One-step model. (*b*) Three-equation model.

the wall, Fig. 5(*a*). Population in the bowl and along the axis is represented by numerous particles of small size and may be related to entities created in the early stages of combustion around the cylinder axis where fuel was available but soot was also oxidized in this high-temperature well-mixed zone. Only a small annular region [Fig. 6(*a*)], close to the bowl edge seems to correlate the presence of soot to oxygen depletion as observed in Fig. 7, where the iso-surface in the bowl embeds the volume where the mass fraction of oxygen is below 6%. The other characteristic population is along the wall and the top of the cylinder with reduced soot oxidation due to heat losses. Their different histories of soot creation dynamics, according to previous explanations, cause soot particles to survive in different regions in the cylinder at the end of the expansion with different implications in soot formation and oxidation, leading to different size distributions and hence the bimodal size distribution found in Fig. 5. However, it should be noted that this distribution is based on primary soot particle size and not on agglomerated clusters effectively gathered at an engine exhaust. The cluster size distribution, not predicted here, may have a different shape.

Conclusion

A soot formation model developed by Tesner et al., already acknowledged for its use in practical applications, has been extended through simple expressions based on kinetic theory. This more comprehensive model is now able to account for an important part of the soot formation history, to a certain level of approximation, which relies on the limiting step of the pyrolysis, the inception based on active radicals from the hydrocarbon breakdown and the growth of the primary particles through collision with each other and the fuel molecules in the surrounding gas. The system of equations is kept as simple as possible to save computational and tuning cost. The intermediate growth specie is bypassed for sake of simplicity, given the numerous approximations encountered at several levels in a real scale 3D simulation. The behavior of this model has been explored with respect to EGR and compared to the one-step model, commonly used by the engine design community. New insights into the soot population history in a diesel engine cylinder have been furthermore gained and active areas related to soot formation have been pointed out.

This modeling, compatible with engine design tools, reveals some important features of the primary particle history in two main zones in the cylinder, leading to two different populations of soot. Near the bowl edge, the remaining soot is explained by oxygen depletion while, close to the top and the wall, heat losses prevent the destruction of the soot so that much particles exist with a large size. The high-temperature and well mixed interior of the cylinder is populated by small soot particles so that two characteristic primary particle sizes may be encountered at the exhaust (this size distribution for the primary soot particles should not prevail of the fractal cluster size released at the exhaust). EGR leads to an increase in soot mass exhaust through an increase in both the number of primary particles and their diameter. The increase in the number of primary particles is related to presence of the fuel for a longer period of time and the increase in diameter may be linked to a weaker oxidation due to lower temperatures in the cylinder.

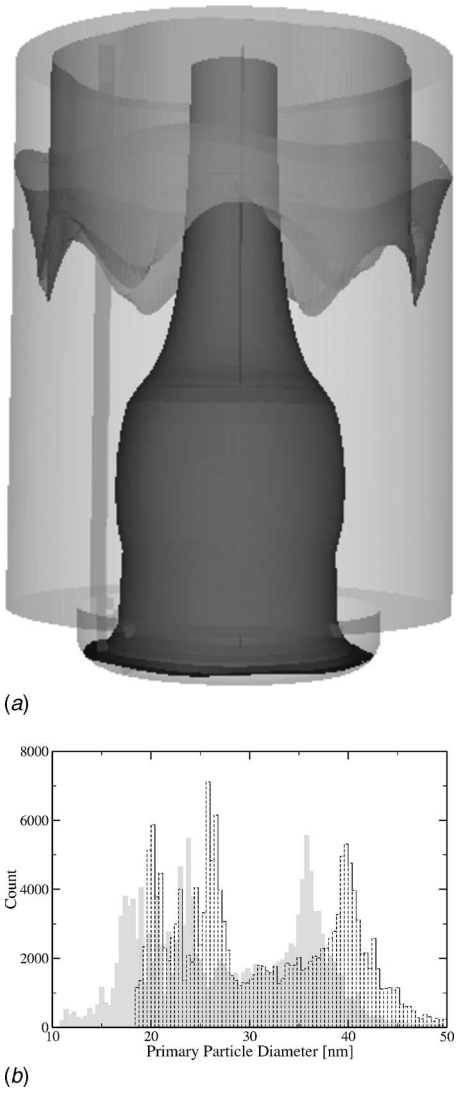
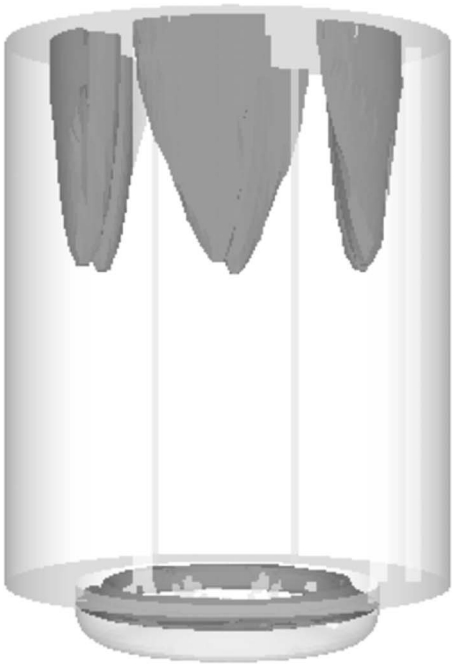


Fig. 5 (*a*) Picture EGR 0%. Plain iso-surface: Diameter below 20 nm. Transparent iso-surface: Diameter above 40 nm. (*b*) Particle diameter distribution at the exhaust (Grey: EGR 0%; Dashed: EGR 8%).

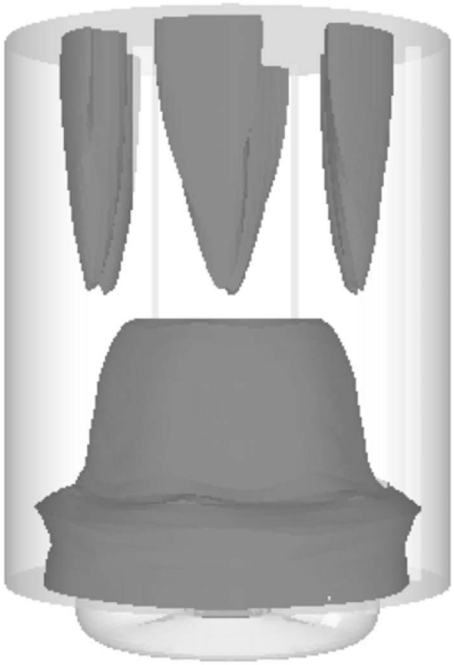
This soot model is of interest due to its ability to predict the primary particle size. However, the soot structure at the exhaust of an engine is an agglomeration of soot primary particles. Those structures appear following a fractal growth process which determines the final weight, size, and number of those clusters. Prediction of agglomeration might be an interesting future development for soot modeling dedicated to diesel engines design. In addition, further validation of the model with experimental data is warranted.

Acknowledgment

This research has been made possible by the funding received from the Government of Canada Program for Energy Research



(*a*)



(*b*)

Fig. 6 Distribution of the two main soot zones at the end of the simulation. (*a*) Iso-surface (8.4×10^{-6}) of the soot mass fraction. (*b*) Iso-surface (1.6×10^{14} kg^{-1}) of the number of primary particles.

and Development (PERD/AFTER). The authors wish to thank Dr. S. C. Kong and Professor R. D. Reitz for providing the grid specification of the engine geometry and parameters.

Nomenclature

Only symbols that are either not available or different from those in the international nomenclature of thermal sciences are listed in the nomenclature.

Fig. 7 Oxygen depletion in the bowl. The iso-surface draws a volume in which oxygen mass fraction is below 6%.

Capital Letters

A_s = primary particle surface (m^2)
D = primary particle diameter (nm)
F = branching-termination coefficient (s^{-1})
K_c = coagulation coefficient (m$^{5/2}$ K$^{-1/2}$ s^{-1})
$K_{G,\text{incep}}$ = growth coefficient (Kg m^3 s^{-1})
N = primary particle number density (m^{-3})
$T_{a_{n_o}}$ = fuel pyrolysis activation temperature (K)

Lowercase Letters

a = soot inception coefficient (s^{-1})
a_0 = preexponential constant of fuel pyrolysis (s^{-1})
b = termination coefficient (m^3/s)
g_0 = termination coefficient (m^3/s)
m_p = nominal mass of primary soot particles (kg)
n = radical number density (m^{-3})
n_o = radical production rate (s^{-1} m^{-3})

Subscripts

F = fuel species
i = cell number
o = inception
s = soot species

References

[1] Belardini, P., Bertoli, C., Ciajolo, A., D'Anna, A., and Del Giacomo, N., 1992, "Three Dimensional Calculations of D. I. Diesel Engine Combustion and Comparison With in Cylinder Sampling Valve Data," SAE Technical Paper No. 92225.
[2] U. S. E. P. A. (EPA), 2002, "Air Quality Criteria for Particulate Matter," Third external review draft EPA/600/P-99/002aC, Research Triangle Park, NC.
[3] 2001, "Réduire les émissions des Véhicules / Vehicle Emission Reductions," 92-82-11363-9, *Proceedings of the European Conference of Ministers of Transport (ECMT)*, OECD, Paris.
[4] Khan, G., and Greeves, A., 1974, *Method for Calculating the Formation and Combustion of Soot in Diesel Engines*, Scripta, Washington, DC, Chap. 25.
[5] Surovikin, V., 1976, "Analytical Description of the Processes of Nucleus-Formation and Growth of Particles of Carbon Black in the Thermal Decomposition of Aromatic Hydrocarbons in the Gas Phase," Solid Fuel Chem., **10**(1), pp. 92–101.
[6] Fusco, A., Knox-Kelecy, A., and Foster, D., 1994, "Application of a Phenomenological Soot Model to Diesel Engine Combustion," Proceedings of the International Symposium COMODIA 94, Yokohama, Japan.
[7] Appel, J., Bockhorn, H., and Frenklach, M., 2000, "Kinetic Modeling of Soot Formation With Detailed Chemistry and Physics: Laminar Premixed Flames of

C_2 Hydrocarbons," Combust. Flame, **121**(1/2), pp. 122–136.
[8] Tao, F., Golovitchev, V., and Chomiak, J., 2004, "A Phenomenological Model for the Prediction of Soot Formation in Diesel Spray Combustion," Combust. Flame, **136**(3), pp. 270–282.
[9] Ahmad, T., Plee, S., and Myers, J., 1985, "Computation of Nitric Oxide and Soot Emissions from Turbulent Diffusion Flames," J. Eng. Gas Turbines Power, **107**(1), pp. 48–53.
[10] Hiroyasu, H., and Nishida, K., 1989, "Simplified Three-Dimensional Modeling of Mixture Formation and Combustion in D. I. Diesel Engine," SAE Technical Paper No. 890269.
[11] Patterson, M., Kong, S., Hampson, G., and Reitz, R., 1994, "Modeling the Effects of Fuel Injection Characteristics on Diesel Engine Soot and NOx Emissions," SAE Technical Paper No. 940523.
[12] Han, Z., and Reitz, R., 1995, "Turbulence Modeling of Internal Combustion Engines using RNG κ-ε Models," Combust. Sci. Technol., **106**(4/6), pp. 267–295.
[13] Kong, S., Han, Z., and Reitz, R., 1995, "The Development and Application of a Diesel Ignition and Combustion Model for Multidimensional Engine Simulation," SAE Technical Paper No. 950278.
[14] Reitz, R., and Rutland, C., 1995, "Development and Testing of Diesel Engine CFD models," Prog. Energy Combust. Sci., **21**(2), pp. 173–196.
[15] Uludogan, A., Xin, J., and Reitz, R., 1996, "Exploring the Use of Multiple Injectors and Split Injection to Reduce D. I. Diesel Engine Emissions," SAE Technical Paper No. 962058.
[16] Han, Z., Uludogan, A., Hampson, G., and Reitz, R., 1996, "Mechanism of Soot and NOx Emission Reduction Using Multiple-Injection in a Diesel Engine," SAE Technical Paper No. 960633.
[17] Chan, M., Das, S., and Reitz, R., 1997, "Modeling Multiple Injection and EGR Effects on Diesel Engine Emissions," SAE Technical Paper No. 972864.
[18] Fuchs, T., and Rutland, C., 1998, "Intake Flow Effects on Combustion and Emissions in a Diesel Engine," SAE Technical Paper No. 980508.
[19] Das, S., Houtz, P., and Reitz, R., 1999, "Effect of Injection Spray Angle and Combustion Chamber Geometry on Engine Performance and Emission Characteristics of a Large Bore Diesel Engine," Proceedings of the Spring Technical Conference, ASME, Vol. 32–1, pp. 1–12.
[20] Jung, D., and Assanis, D., 2001, "Multi-Zone DI Diesel Spray Combustion Model for Cycle Simulation Studies of Engine Performance and Emissions," SAE Technical Paper No. 2001-01-1246.
[21] Tesner, P., Snegiriova, T., and Knorre, V., 1971, "Kinetics of Dispersed Carbon Formation," Combust. Flame, **17**(2), pp. 253–260.
[22] Mehler, R., Amann, M., and Schopp, W., 2002, "A Methodology to Estimate Changes in Statistical Life Expectancy Due to the Control of Particulate Matter Air Pollution," Technical Report No. IR-02-035, International Institute for Applied Systems Analysis, Laxenburg, Austria.
[23] Parker, J., 2005, "Air Pollution and Birth Weight among Term Infants in California," Pediatrics, **115**(1), pp. 121–128.
[24] Lee, J., Goto, Y., and Okada, M., 2002, "Measurement of the Diesel Exhaust Particle Reduction Effect and Particle Size Distribution in a Transient Cycle Mode With an Installed Diesel Particulate Filter (DPF)," SAE Technical Paper No. 2002-01-1005.
[25] Moss, J., Stewart, C., and Young, K., 1995, "Modeling Soot Formation and Burnout in a High Temperature Laminar Diffusion Flame Burning under Oxygen-Enriched Conditions," Combust. Flame, **101**(4), pp. 491–500.
[26] Leung, K., Lindstedt, R., and Jones, W., 1991, "A Simplified Reaction Mechanism for Soot Formation in Nonpremixed Flames," Combust. Flame, **87**(3/4), pp. 289–305.
[27] Bédat, B., Egolfopoulos, F. N., and Poinsot, T., 1999, "Direct Numerical Simulation of Heat Release and NO_x Formation in Turbulent Nonpremixed Flames," Combust. Flame, **119**(1/2), pp. 69–83.
[28] Nagle, J., and Strickland-Constable, R., 1962, "Oxidation of Carbon Between 1000–2000°C," *Proceedings of the Fifth Conference on Carbon*, Pergamon, London, pp. 154–164.
[29] Fenimore, C., and Jones, G., 1967, "Oxidation of Soot by Hydroxyl Radicals," J. Phys. Chem., **71**(3), pp. 593–597.
[30] Bradley, D., Dixon-Lewis, G., El-Din Habik, S., and Mushi, E., 1984, "The Oxidation of Graphite Powder in Flame Reaction Zones," *Proceedings of the 20th International Symposium on Combustion*, The Combustion Institute, Pittsburgh, pp. 931–940.
[31] Garo, A., 2004, private communication.
[32] Westbrook, C., and Dryer, F., 1984, "Chemical Kinetic Modelling of Hydrocarbon Combustion," Prog. Energy Combust. Sci., **10**(1), pp. 1–57.
[33] Magnussen, B., and Hjertager, B., 1976, "On Mathematical Modeling of Turbulent Combustion With Special Emphasis on Soot Formation and Combustion," *Proceedings of the 16th International Symposium on Combustion*, The Combustion Institute, Pittsburgh, pp. 719–729.
[34] Methodology volume, Star-CD, 2004, Edition.
[35] Huh, K., and Gosman, A., 1991, "A Phenomenological Model of Diesel Spray Atomisation," Proceedings of the International Conference on Multiphase Flows (ICMF'91), Tsukuba, Japan.
[36] Reitz, R., and Diwakar, R., 1986, "Effect of Drop Breakup on Fuel Sprays," SAE Technical Paper No. 860469.
[37] Mastorakos, E., and Wright, Y. M., 2003, "Simulations of Turbulent Spray Auto-Ignition With Elliptic Conditional Moment Closure," Proceedings of the European Combustion Meeting, Orléans, France.
[38] Yakhot, V., Orszag, S., Thangam, S., Gatski, T., and Speziale, C., 1992, "Development of Turbulence Models for Shear Flows by a Double Expansion Technique," Phys. Fluids A, **4**(7), pp. 1510–1520.
[39] Launder, B., and Spalding, D., 1974, "The Numerical Computation of Turbulent Flows," Comput. Methods Appl. Mech. Eng., **3**(2), pp. 269–289.
[40] Jayatilleka, C., 1969, "The Influence of Prandtl Number and Surface Roughness on the Resistance of the Laminar Sub-Layer to Momentum and Heat Transfer," Prog. Heat Mass Transfer, **1**, pp. 193–330.
[41] McLandress, A., Emerson, R., McDowell, P., and Rutland, C., 1996, "Intake and in-Cylinder Flow Modeling Characterization of Mixing and Comparison With Flow Bench Results," SAE Technical Paper No. 960635.
[42] Neill, W. S., and Chippior, W., 2002, NRC Report No. PET-1528-02S.
[43] Hergart, C., Barths, H., and Peters, N., 2000, "Using Representative Interactive Flamelets in Three-Dimensional Modelling of the Diesel Combustion Process Including Effects of Heat Transfer," Proceedings of the International Multidimensional Engine Modeling User's Group Meeting, Madison, WI.
[44] Hergart, C., and Peters, N., 2002, "Applying the Representative Interactive Flamelet Model to Evaluate the Potential Effect of Wall Heat Transfer on Soot Emissions in a Small-Bore Direct-Injection Diesel Engine," J. Eng. Gas Turbines Power, **124**(4), pp. 1042–1052.
[45] Tao, F., Srinivas, S., Reitz, R. D., and Foster, D. E., 2004, "Current Status of Soot Modeling Applied to Diesel Combustion Simulations," Proceedings of the International Symposium COMODIA 2004, Yokohama, Japan.
[46] Khan, I. M., Wang, C. H. T., and Langridge, B. E., 1971, "Coagulation and Combustion of Soot Particles in Diesel Engines," Combust. Flame, **17**(3), pp. 408–419.
[47] Smith, O. I., 1981, "Fundamentals of Soot Formation in Flames With Application to Diesel Engine Particulate Emissions," Prog. Energy Combust. Sci., **7**(4), pp. 275–291.

Journal of Engineering for Gas Turbines and Power

Technical Brief

Assessment of the Computing Time for the IAPWS-IF97 Equations

Kiyoshi Miyagawa
4-12-11-628 Nishiogu, Arakawa-ku,
Tokyo 116-0011, Japan
e-mail: miyagawa.kiyoshi@nifty.ne.jp

Computing times of equations based on the IAPWS Industrial Formulation 1997 for the Thermodynamic Properties of Water and Steam (IAPWS-IF97) were tested. Modern computer systems are optimized for "simple" computational operations, which favors the simple structure of IAPWS-IF97. Provision of "backward equations," which are approximation of inverse equations, is one of the features of IAPWS-IF97. The backward equations showed much shorter computing times than iterative routines, which had been used to calculate with several independent variables. IAPWS-IF97 is faster than the equations of IAPWS Formulation 1995 for the Thermodynamic properties of Ordinary Water Substance for General and Scientific Use (IAPWS-95) by factors 70 to 200 times. [DOI: 10.1115/1.2436576]

1 Introduction

In 1997, the International Association for the Properties of Water and Steam (IAPWS) adopted a new formulation for the thermodynamic properties of water and steam for industrial use. This formulation, called IAPWS Industrial Formulation 1997 for the Thermodynamic Properties of Water and Steam (IAPWS-IF97), formed the basis for power plant calculations and other applications in energy engineering. The key motivation behind the development of IAPWS-IF97 was to improve the computing speed.

The main equations of IAPWS-IF97 are specified in the IAPWS release [1]. Figure 1 shows the applicable temperature–pressure ranges covered by the IAPWS-IF97 equations. The main part of the formulation is defined for temperature range of 273.15–1073.15 K and pressure range up to 100 MPa. This is the most important range of computation in steam power engineering. It is split into three regions: Region 1 is the compressed liquid region for which a Gibbs free energy $g_1(p,T)$ is given as the basic equation; Region 2 is the superheated vapor region for which a different Gibbs free energy $g_2(p,T)$ is provided; and Region 3 is the supercritical region for which a Helmholtz free energy $f_3(v,T)$ is provided. The forward equations of thermodynamic properties, $z(p,T)$ for Regions 1 and 2 or $z(v,T)$ for Region 3, are derived from the basic equations.

Submitted to ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received August 17, 2006; final manuscript received October 29, 2006. Review conducted by Dilip R. Ballal.

Between 2001 and 2005, IAPWS adopted several supplementary releases for the so-called "backward equations" [2–5]. The backward equations are explicit expressions of functions for thermodynamic properties and functions for region test with respect to main variable set, other than that for the forward equations, such as $z(p,h)$ and $z(p,s)$. The backward functions are approximations of the exact inverse functions of the forward equations. They are numerically inconsistent with the exact inverse functions. However, the deviation is quite small for typical engineering applications. The main advantage is a significantly shorter computing time. In contrast, solving the exact inverse functions requires very time-consuming iterative routines.

To shorten computing time, the equations of IAPWS-IF97 were designed to be "simple" polynomials expressed as a series of additions and multiplications without the use of exponential functions.

Detailed information on the development of the IAPWS-IF97 equations are seen in background papers by the developers [6,7].

The IAPWS has tested the computing times of the equations in the releases at the development stage using the latest computing environments at the time. The equations in the original IAPWS-IF97 release were tested on a 16-bit DOS PC with 486 DX/33 MHz CPU. The more recent supplementary releases were tested on 32-bit Windows PCs with various classes of Pentium. It is difficult to compare past test results because the differences in computing times on different testing environments were very large.

The author reevaluated them on the modern PC environment and compared item with several reference routines. Computing times were measured on a Windows-XP PC running Pentium 4/3.0 GHz. In addition, the author performed the tests in other modern environments, which will be discussed later.

2 Reference Codes

To evaluate the computing speed of the IAPWS-IF97 routines, several reference codes were used.

2.1 IFC-67 ASME Codes. At the development stage of IAPWS-IF97, it was required that its computing speed should be much faster than those of the previous international formulation IFC-67 [8]. In the evaluation stage of IAPWS-IF97, special codes based on IFC-67 were used as the references. The reference codes were taken from an ASME program package [9], which had been widely used in the industrial applications.

To compare the computing speed of IAPWS-IF97 equations with that of IFC-67, IAPWS specified official weighting factors based on the frequency of use in industrial applications, which were obtained by an international survey [6].

2.2 Iterative Routines Using IAPWS-IF97 Forward Functions. The motivation for the development of the IAPWS-IF97 backward equations was to avoid time-consuming iterative routines required by the forward equations. IAPWS has developed reference iterative codes to evaluate the backward equations in the releases. Two-dimensional Newton's, secant, or Brent's [10]

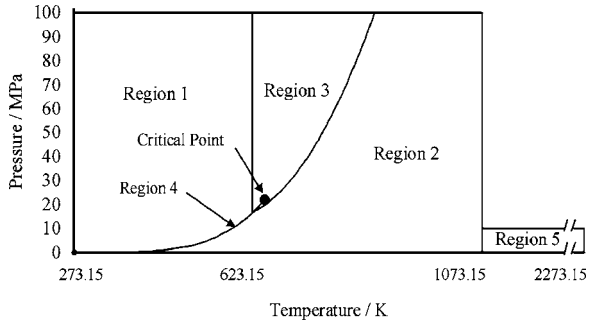


Fig. 1 Structure of IAPWS-IF97

method were used as the root-finding algorithms.

The convergence criteria of the iterative reference codes were selected so that the accuracies are similar to those of the backward functions. The reference codes were programmed with a minimum number of starting points to allow the routines to converge.

2.3 Iterative Routines Using IAPWS-95 Functions. The IAPWS-IF97 was developed as an approximation of the IAPWS Formulation 1995 for the Thermodynamic Properties of Ordinary Water Substance for General and Scientific Use (IAPWS-95) [11,12]. The basic equation of IAPWS-95 is a Helmholtz free energy $f(T,\rho)$, which covers a wider range than that of IAPWS-IF97 with a single equation. Because of its complicated equation form, its computing speed is much slower than that of IAPWS-IF97. Since information about the computing speed of IAPWS-95 is not well known, the author tested it and compared it with that of IAPWS-IF97.

3 Computing Times for IAPWS-IF97 and IFC-67

Computing times for IFC-67 and IAPWS-IF97 were tested by IAPWS in 1996 on a 16-bit DOS PC with 486 DX/33 MHz CPU. The average computing time ratio (CTR) was 5.1 based on the official weighting factors [6]. The author performed a similar test on a Windows-XP PC running Pentium 4/3.0 GHz and the average CTR was 8.3. The higher CTR means that modern environments, such as Pentium 4 with Windows, are optimized for "simple" computational operations and therefore favor the simpler structure of IAPWS-IF97.

4 Computing Times for IAPWS-IF97 Backward Equations and Reference Iterative Routines

Tables 1 and 2, respectively, show the computing times of the IAPWS-IF97 backward equations for thermodynamic properties and the reference iterative routines with similar accuracy level.

Table 1 Computing times for backward equations and iterative routines for thermodynamic properties

Function	*A*: backward equation (μs/call)	*B*: reference iterative code (μs/call)	CTR *B*/*A*
$T_1(p,h)$	0.050	1.26	25
$T_1(p,s)$	0.052	1.96	38
$p_1,T_1(h,s)$	0.117	4.14	35
$T_2(p,h)$	0.117	1.29	11
$T_2(p,s)$	0.166	2.31	14
$p_2,T_2(h,s)$	0.248	11.51	46
$v_3(p,T)$	0.278	4.64	17
$v_3,T_3(p,h)$	0.366	5.08	14
$v_3,T_3(p,s)$	0.345	4.80	14
$v_3,T_3(h,s)$	0.517	4.98	10
$T_{sat}(h,s)$	0.109	8.41	77

Table 2 Computing times for backward equations and iterative routines for region tests

Function	*A*: backward equation (μs/call)	*B*: reference iterative code (μs/call)	CTR *B*/*A*
$p_{sat}(h)$	0.072	2.71	31
$p_{sat}(s)$	0.051	2.54	50
$h_{b13}(s)$	0.113	2.17	19
$I_{b23}(h,s)$	0.361	2.67	7
$h_1'(s)$	0.100	3.91	39
$h_3'(s)$	0.077	28.0	365
$h_2''(s)$	0.237	5.20	22
$h_3''(s)$	0.079	29.1	368

The backward equations were much faster than the iterative routines. In particular, very high CTR values were obtained in backward equations for the region tests related to Region 3.

Although the accuracies of backward equations are enough for the usual applications, users might require inverse equations, which have very high numerical consistency with the forward equations. In such cases, iterative routines must be used with reduced convergence criteria.

The author tested the iterative routines in Table 1 with convergence criteria reduced to 1/1000. In addition, the author tested the same thing using IAPWS-IF97 backward equations as the starting point of the iteration. Table 3 shows the test results. The starting points from backward equations are useful in reducing the computing time of the iteration.

5 Computing Times Tested on Different Computers

The author tested the computing times in several different computer environments. Table 4 shows the arithmetic average of computing times for the IAPWS-IF97 forward and backward functions. Table 4 also contains the adjusted time defined as the product of the time and the clock frequency of CPU. The adjusted time values for the Pentium 4 CPUs are about 6 and higher (slower) than others. This suggests that CPUs other than Pentium 4 are better optimized for floating point operations.

Athlon 64 3500+ was the fastest among the tested environments. There was no significant difference between Windows XP and Me.

6 Computing Times for IAPWS-95

Table 5 shows computing times for iterative routines of IAPWS-IF97 and IAPWS-95. The values of CTR are about 70–

Table 3 Computing times for iterative routines with reduced convergence criteria to 1/1000 of those in Table 1

	Starting point		
Function	*A*: IAPWS-IF97 backward equations (μs/call)	*B*: Same as Table 6 (μs/call)	CTR *B*/*A*
$T_1(p,h)$	0.70	1.58	2.3
$T_1(p,s)$	0.95	2.22	2.3
$p_1,T_1(h,s)$	2.08	4.93	2.4
$T_2(p,h)$	0.75	1.58	2.1
$T_2(p,s)$	1.28	2.72	2.1
$p_2,T_2(h,s)$	2.39	12.7	5.3
$v_3(p,T)$	2.75	5.37	1.9
$v_3,T_3(p,h)$	3.00	5.74	1.9
$v_3,T_3(p,s)$	3.16	5.83	1.8
$v_3,T_3(h,s)$	1.87	4.91	2.6
$T_{sat}(h,s)$	5.22	9.90	1.9

Table 4 Arithmetic average of computing times for IAPWS-IF97 equations on different PCs

CPU	A: clock (GHz)	OS	B: time (μs/call)	Adjusted time $A \times B$
Pentium 4	3.0	Win XP	0.199	0.60
Pentium 4	2.53	Win XP	0.232	0.59
Pentium 4	1.5	Win XP	0.396	0.59
Pentium M	1.1	Win XP	0.327	0.36
Pentium III	1.2	Win XP	0.258	0.31
Pentium III	1.2	Win ME	0.255	0.31
Athlon 64 3500+	2.2	Win XP	0.130	0.29

200. Although IAPWS-95 has several advantages, such as single basic equation and wider applicable range, it is markedly slower than IAPWS-IF97. The users should consider whether IAPWS-95 should be used in their applications with sacrifice in computing time.

7 Conclusions

Computing times for IAPWS-IF97 equations were determined. IAPWS-IF97 equations were much faster than the previous international formulation IFC-67 equations. Modern computer systems are optimized for "simple" computational operations and therefore favor the simpler structure of IAPWS-IF97.

The backward equations of IAPWS-IF97 are much faster than the iterative process and they are useful in reducing the computing time of applications. Users who need iterative routines for higher numerical consistency should use the backward equations for the initial guess to achieve shorter computing time.

IAPWS-IF97 equations are faster than those of IAPWS-95 by factors of about 70–200. The users should consider whether IAPWS-95 should be used in their applications with sacrifice in computing time.

Acknowledgment

The assessment was initiated by a suggestion of Hans-Joachim Kretzschmar and Katja Knobloch. The author is grateful for their helpful advice on the assessment. Information about the IAPWS Releases, Supplementary Releases can be obtained from the IAPWS website: www.iapws.org. Further information about the computing time tests can be obtained from the author.

Table 5 Computing times for iterative process using forward equations $z(v,T)$ of IAPWS-IF97 region 3 and IAPWS-95

	Forward equation		
Function	A: IAPWS-IF97 (μs/call)	B: IAPWS-95 (μs/call)	CTR B/A
$v,T(p,h)$	5.08	1014	200
$v,T(p,s)$	4.80	1003	209
$v,T(h,s)$	4.98	1036	208
$v(p,T)$	4.64	309	67

Nomenclature

Variables

f = Helmholtz free energy
g = Gibbs free energy
h = specific enthalpy
p = pressure
s = specific entropy
T = temperature
v = specific volume
z = general properties
ρ = density
CTR = computing time ratio

Superscripts

$'$ = saturation liquid state
$''$ = saturation vapor state

Subscripts

1,2,3 = regions of IAPWS-IF97
sat = saturation state
$b13$ = region boundary 1-3
$b23$ = region boundary 2-3

References

[1] IAPWS, 1997, *Release on the IAPWS Industrial Formulation 1997 for the Thermodynamic Properties of Water and Steam*, available at http://www.iapws.org.

[2] IAPWS, 2001, *Supplementary Release on Backward Equations for Pressure as a Function of Enthalpy and Entropy $p(h,s)$ to the IAPWS Industrial Formulation 1997 for the Thermodynamic Properties of Water and Steam*, available at http://www.iapws.org.

[3] IAPWS, 2004, *Revised Supplementary Release on Backward Equations for the Functions $T(p,h)$, $v(p,h)$ and $T(p,s)$, $v(p,s)$ for Region 3 of the IAPWS Industrial Formulation 1997 for the Thermodynamic Properties of Water and Steam*, available at http://www.iapws.org.

[4] IAPWS, 2004, *Supplementary Release on Backward Equations $p(h,s)$ for Region 3, Equations as a Function of h and s for the Region Boundaries, and an Equation $T_{sat}(h,s)$ for Region 4 of the IAPWS Industrial Formulation 1997 for the Thermodynamic Properties of Water and Steam*, available at http://www.iapws.org.

[5] IAPWS, 2005, *Supplementary Release on Backward Equations for Specific Volume as a Function of Pressure and Temperature $v(p,T)$ for Region 3 of the IAPWS Industrial Formulation 1997 for the Thermodynamic Properties of Water and Steam*, available at http://www.iapws.org.

[6] Wagner, W. et al., 2000, "The IAPWS Industrial Formulation 1997 for the Thermodynamic Properties of Water and Steam," J. Eng. Gas Turbines Power, **122**, pp. 150–182.

[7] Kretzschmar, H.-J., et al., 2004, "Supplementary Backward Equations for the Industrial Formulation IAPWS-IF97 of Water and Steam for Fast Calculations of Heat Cycles, Boilers, and Steam Turbines," *Water, Steam, and Aqueous Solutions for Electric Power, Proceedings of the 14th International Conference on the Properties of Water and Steam*, Maruzen Co., Ltd., Tokyo, Japan, pp. 34–45.

[8] Meyer, C. A., et al., 1993, *ASME Steam Tables*, 6th ed., ASME, New York, pp. 19–40.

[9] Meyer, C. A., et al., 1993, *ASME Steam Tables*, 6th ed., ASME, New York.

[10] Press, H. W., et al., 1992, *Numerical Recipes in Fortran*, 2nd ed., Cambridge University Press, New York.

[11] IAPWS, 1996, *Release on the IAPWS Formulation 1995 for the Thermodynamic Properties of Ordinary Water Substance for General and Scientific Use*, available at http://www.iapws.org.

[12] Wagner, W. et al., 2002, "The IAPWS Formulation 1995 for the Thermodynamic Properties of Ordinary Water Substance for General and Scientific Use," J. Phys. Chem. Ref. Data, **31**, pp. 387–535.

Journal of Engineering for Gas Turbines and Power

Book Review

Rotordynamics, by Agnieszka (Agnes) Muszynska. CRC Press, Taylor & Francis Group, Boca Raton, FL, 2005, 1128 pp., ISBN: 0-8247-2399-6

REVIEWED BY JERZY T. SAWICKI[1]

This 1128-page monograph presents comprehensive rotordynamic problems in rotating machinery. These problems are recognized by machine monitoring data and correlated with modal models obtained by multi-mode modal identification. Being directly related to the modal characteristics, with practically identifiable parameters, multi-mode models have a solid base for further application in such areas as stability and post-stability self-excited vibrations, sensitivity analysis, active control, solid-fluid interaction, local/global dynamic effects, fractional resonances, chaotic vibrations, and many others. The emphasis is on understanding rotordynamic physical phenomena. Multitudes of experiments, described in detail, not only support the presented analytical material, but perhaps may also stimulate an enthusiasm to repeat and extend these experiments. This book also sets out fundamentals of vibration monitoring and malfunction diagnosis of rotating machines.

In Chapter 1, dynamics of the fundamental two-lateral-mode isotropic model of a rotor, as well as principles of modal identification are introduced. This model is extended in Chapter 3. Chapter 2 presents an introduction to vibration monitoring and data processing for rotating machinery, and reviews trends in machinery management and monitoring programs.

The most comprehensive Chapter 4 discusses important fluid-related problems known as "rotor fluid-induced instabilities," arising from rotor/fixed structure clearances in rotating machines. The Bently/Muszynska (B/M) model of the fluid forces in rotor clearances, based on the strength of the fluid circumferential flow, represented by the "fluid circumferential average velocity ratio" (denoted by λ is introduced. The rotor malfunction phenomena, fluid whirl, and fluid whip, are identified as limit cycles of self-excited vibrations. New phenomena, such as higher mode fluid whirl and fluid whip, are discussed, experimentally demonstrated, and adequately modeled. This chapter also introduces a formal, but unconventional, derivation of the fluid forces, starting from the classical Reynold's equation. The presentation clarifies some paradoxes, which are described in the rotordynamic literature. Comparisons are also presented between the B/M model and classical bearing/seal coefficients, thus providing a tool for extracting the meaning value of λ and other parameters of the B/M model from the coefficient tables. Identification methods of rotor/bearing/support system parameters are also discussed.

Chapter 5 presents rotor-to-stator dry contact-related rubbing. Experimental results, mathematical models of rubbing rotors, their solutions, and diagnostic recognition patterns are presented.

Chapter 6 discusses a series of important rotordynamic topics: balancing, stress in rotating and laterally vibrating rotors, multimode modal modeling, basing on modal data, rotor coupled lateral/torsional vibrations, loose rotating parts, role of damping in rotating structures, and early detection of rotor cracks using vibration data.

Chapter 7 outlines vibration diagnosis of various malfunctions in rotating machines based on vibrational data and illustrated by means of basic simplified mathematical rotor models.

Extensive references and a list of mathematical notations follow each chapter. Ten Appendices and a Glossary of terminology complete the book.

The book does not contain worked examples and there are no problems or suggested laboratory exercises at the end of each chapter that could be assigned to students. This would make it difficult to use as a textbook in a course dealing with dynamics and diagnostics of rotating machinery. This monograph represents an invaluable resource for professionals and practicing engineers concerned with rotating machines.

[1]Department of Mechanical Engineering, Cleveland State University, Cleveland, OH 44115-2214; e-mail: j.sawicki@csuohio.edu

