Transactions of the ASME®

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).

INDEXED by Applied Mechanics Reviews and Engineering Information, Inc. Canadian Goods & Services Tax Registration #126148048

Journal of Engineering for Gas Turbines and Power

Published Quarterly by The American Society of Mechanical Engineers

VOLUME 123 • NUMBER 2 • APRIL 2001

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.

Journal of Engineering for Gas Turbines and Power APRIL 2001

Volume 123, Number 2

Thermodynamic Analysis of Air-Cooled Gas Turbine Plants

E. A. Khodak

G. A. Romakhova

St.-Petersburg State Technical University,
St.-Petersburg, Russia

At present high temperature, internally cooled gas turbines form the basis for the development of highly efficient plants for utility and industrial markets. Minimizing irreversibility of processes in all components of a gas turbine plant leads to greater plant efficiency. Turbine cooling, like all real processes, is an irreversible process and results in lost opportunity for producing work. Traditional tools based on the first and second laws of thermodynamics enable performance parameters of a plant to be evaluated, but they give no way of separating the losses due to cooling from the overall losses. This limitation arises from the fact that the two processes, expansion and cooling, go on simultaneously in the turbine. Part of the cooling losses are conventionally attributed to the turbine losses. This study was intended for the direct determination of lost work due to cooling. To this end, a cooled gas turbine plant has been treated as a work-producing thermodynamic system consisting of two systems that exchange heat with one another. The concepts of availability and exergy have been used in the analysis of such a system. The proposed approach is applicable to gas turbines with various types of cooling: open-air, closed-steam, and open-steam cooling. The open-air cooling technology has found the most wide application in current gas turbines. Using this type of cooling as an example, the potential of the developed method is shown. Losses and destructions of exergy in the conversion of the fuel exergy into work are illustrated by the exergy flow diagram.
[DOI: 10.1115/1.1341204]

1 Introduction

Today, efficiency levels for new large natural-gas-fired gas turbine plants are in the range of 34–36 percent. Gas turbine based combined plants achieved efficiencies of 54–56 percent. The use of advanced cooling techniques and improvements in materials provide the opportunity for increasing the turbine inlet temperature and for improving plants efficiencies. This trend will continue.

There are two opposing effects, if the turbine inlet temperature is raised without improving a cooling technology. Raising the turbine inlet temperature tends to the higher mean temperature of heat supply and to increase the plant efficiency. But with an increase in turbine inlet temperature there is an increase in heat flows from the hot gases to the coolant, if the mean blade temperature is kept on. This means increasing the required coolant flow. The transfer of heat over a finite temperature difference between the gas and coolant flows represents a lost opportunity for producing work. The irreversibility occurs in each process through which the coolant goes, and represents, also, a lost opportunity for producing work. Thus raising the turbine inlet temperature tends to increase irreversibility associated with cooling. A gain in efficiency due to an increase in the mean temperature of heat supply may sometimes be partly or more than offset by the effect of resulting greater irreversibility.

The amount of lost work due to cooling is greatly dependent on cooling effectiveness falling drastically with an improvement in cooling technology. This would result in a higher plant efficiency additional to that resulting from the increase in the turbine inlet temperature.

In defining ''cooling losses,'' it is first necessary to establish the type and the location of these losses. Conventional air cooling is now widely used in gas turbines, and this is the type of cooling considered in this paper. Cooled and uncooled gas turbine plants differ in open circuits on which they operate. The Joule–Brayton cycle cannot be said to be the basis of this plant. The fuel energy is supplied to the part of the compressed air mass flow rate only. The remainder of the air mass flow is extracted from the compressor, ducted to the turbine blades, and used as a coolant. Absorbing the heat extracted through wall and the heat extracted from the fluid stream as the result of mixing cooling flows produce work during the expansion in the turbine. Cooling losses derive from the introduction of supplementary irreversibilities associated with the modification of the open circuit to the turbine cool. The determination of such losses is a matter for thermodynamic analysis.

2 Methodology

An air-cooled gas turbine power plant is an open-circuit steady-flow device which converts energy of fuel into mechanical work. A distinguishing feature of recent turbines is that large cooling air usage is required for good cooling.

An established open-air cooling system for a turbine incorporates a rotor cooling circuit and several stationary cooling circuits provided by the air extracted from the compressor discharge and from interstate compressor bleeds. If the temperature of the bled air is too high to be used directly to cool turbine hot components, the air flows may be precooled in external coolers. Currently highly complex cooling schemes are used to cool vanes and blades. When passing through cooling channels the cooling air absorbs the heat extracted from the hot gases. The heated coolant is ejected into blade path through exit holes in aerofoils, trailing edges, and leading edges and is mixed with the main flow. The compressor bleed air is also supplied to interstage disk cavities to cool disk and to interstage seals to preclude the injection of hot gases.

Thus the expansion in the air-cooled turbine is accompanied by nearly-continuous mixing with cooling air. As the working fluid temperature drops, the intensity of the cooling air supply decreases. The flow through the turbine is no longer treated as being adiabatic. An air-cooled turbine is an open thermodynamic system that exchanges both energy and mass with the exterior. Figure 1 shows the simplest arrangement of a plant with cooling air supplied ''continuously'' to the turbine.

Contributed by the Advanced Energy Systems of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the AES Division, July 14, 2000; final revision received by the ASME Headquarters Aug. 1, 2000. Editor: H. D. Nelson.

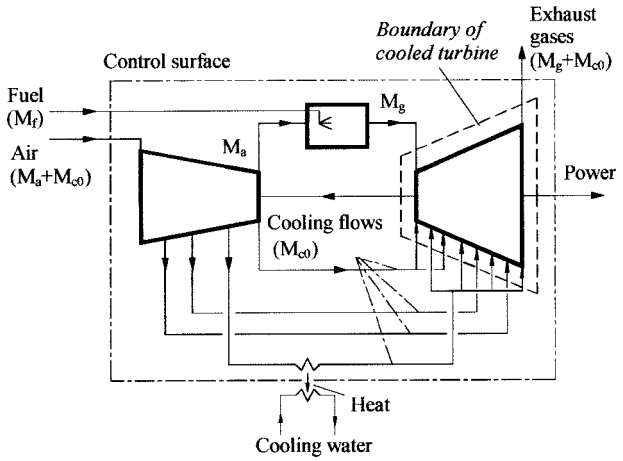


Fig. 1 Gas turbine plant with continuous air supply for cooling

The two open circuits together form an air-cooled gas turbine plant. The air entering the combustion chamber after leaving the compressor (M_a) and the products of combustion ($M_g = M_a + M_f$, where M_f is the mass rate of the fuel supplied) pass through one of them (see Fig. 1). The other circuit is the coolant circuit through which the cooling air flows (M_{c0}) pass. The essential difference of the first circuit from the conventional circuit on which an uncooled gas turbine plant operates is that the expansion of the products of combustion in the turbine is accompanied by heat removing. This heat is supplied to the coolant circuit.

An air-cooled gas turbine plant can be treated as a binary plant, in which a combination of working substances is used. The upper plant is supplied with reactants (fuel and air), produces power, rejects heat during the expansion in the turbine, and discharges the products of combustion. This plant acts as a topping plant to a bottoming plant using the heat rejected as a heat source. The air bled from the compressor to cool gas turbine components (the coolant) is a working substance for the bottoming plant. In an air-cooled gas turbine plant a topping open circuit is superposed directly on a bottoming circuit.

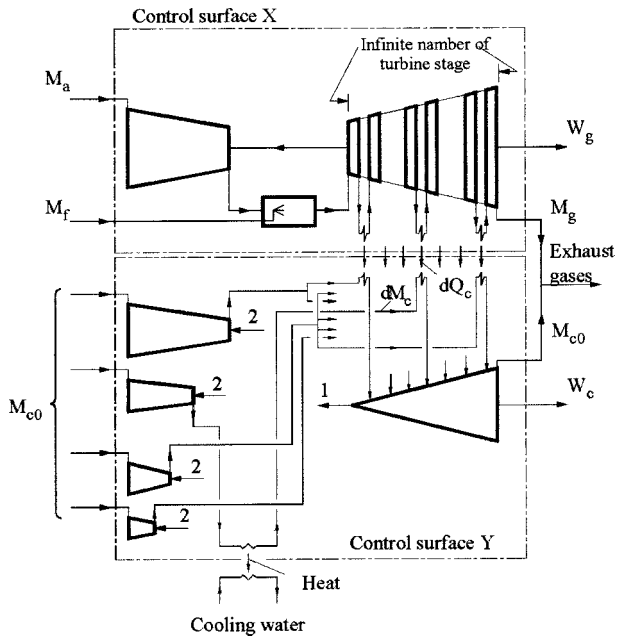


Fig. 2 Topping and bottoming circuits. 1—power to drive compressors, 2—work input to compressors

Figure 2 shows an interpretation of how an air-cooled gas turbine plant can perform as a hypothetical ''combined'' plant. The plant contained within control surface X comprises the topping plant and that within control surface Y comprises the bottoming plant.

The air entering the plant is split before compression into two streams (M_a) and (M_{c0}). After compression the first stream enters the combustion chamber, the flue gases from which supply the working fluid (the gas) for the gas turbine in the topping plant.

The second stream (M_{c0}) is split into infinite number of flows after compression. They enter an infinite number of coils, being placed between an infinite number of turbine stages, as shown in Fig. 2. The pressure drop across each stage is infinitely small. The velocities of all streams are so small that kinetic energies may be ignored. An elementary quantity of air (dM_c) enters the coil at some temperature below the temperature of the incoming gas stream. When passing through a coil it absorbs the heat (dQ_c), and its temperature increases only an infinitesimal amount less than the leaving gas stream. The two fluid streams, gas and coolant, leaving a coil are at the same pressure, but they do not mix, as in the case of a real plant. The heat transfer from the gas to the coolant takes place across a variable finite temperature difference and so is irreversible. Heated coolant expands through its own work-producing turbine. Part of the turbine work developed is used to drive compressors.

The expansion of the gas in its own turbine is accompanied by heat transfer and the expansion of the coolant in its own turbine takes place in the presence of energy transfer, so both processes are not adiabatic. The pressure and temperature drops are the same across each turbine.

The extraction of heat (Q_c) causes a reduction in the work done by the gas, but this reduction is partially offset by the work output from the bottoming plant. When $Q_c = 0$, ''the combined plant'' becomes a conventional uncooled gas turbine plant.

The quantity of heat extracted (Q_c) meets the sum of the heat removed from the gas through walls of blades, and that removed by mixing the gas and the cooling air flows. The former is much less than the latter (Table 1). It represents the heat that must be extracted through wall to achieve required metal temperature and is determined by the external heat transfer process in turbine blades. The latter depends on the required cooling air flow rate, and so varies with cooling effectiveness.

Thus an air-cooled gas turbine plant may be treated as a hypothetical combined plant, the upper part operating on an open circuit, producing work (W_g), rejecting the products of combustion (M_g) and heat (Q_c). This heat is supplied to the lower part of the plant producing work (W_c) and rejecting the air mass flow (M_{c0}). In terms of thermodynamic systems, an air-cooled gas turbine plant represents a compound system comprising two systems that exchange heat with one another. Let us call the former system the gas system and the letter system the coolant system.

Compared to an uncooled gas turbine plant, there are three features that result in additional lost opportunities for producing work:

1 the irreversibility involved in transfer of heat between the gas and the coolant flows (between the topping and the bottoming plants, or in other words, between the gas and the coolant systems);

2 irreversibilities involved in all processes that make up the coolant circuit (irreversibilities within the bottoming plant);

3 the irreversibility of the mixing process of two streams different in chemical composition.

Since the lost work due to irreversibility is equal to the increase in the anergy (energy being unavailable for work producing), so only the concepts of thermodynamic availability and exery provide a possibility of determining ''the cooling losses.'' Exergy principles and their applications are described by Szargut et al.

[2], Kotas [3], and Moran and Sciubba [4]. The application of the concept of exergy to an air-cooled gas turbine plant is illustrated in Sections 3 and 4.

Having determined the temperatures and pressures at all points of both circuits, as well as determined the quantity of heat removed from the gas, conventional criteria of thermodynamic performance may be used to illustrate how much work is done by the gas and coolant.

The overall efficiency of the plant (η) may be written in terms of the overall efficiencies (η_g) and (η_c) of the hypothetical topping and bottoming plants:

$$\eta = \frac{W}{F} = \frac{W_g + W_c}{F} = \frac{W_g}{F} + \frac{Q_c}{F}\frac{W_c}{Q_c} = \eta_g + \chi_c \eta_c , \qquad (1)$$

where F is the fuel energy supplied and $\chi_c = Q_c/F$ is the ratio of the heat extracted from the topping plant to the fuel energy supplied.

This elementary equation may be illustrated for a sample gas turbine plant. Details of the plant are given in Section 4. The plant develops 167.4 MW. The efficiency of the plant, with the heat supplied at 470.6 MW (based on the lower calorific value of the fuel), is 0.356. The rate of work output from the topping plant is 154.2 MW with a heat removed during expansion of 72.7 MW. The bottoming plant is supplied with this heat and develops 13.2 MW.

The efficiencies of the topping and bottoming plants are $\eta_g = 154.2/470.6 = 0.328$ and $\eta_c = 13.2/72.7 = 0.182$, respectively. About 15.5 percent ($\chi_c = 72.4/470.6 = 0.155$) of the fuel energy is not converted into work by the topping plant, but is removed during the expansion and is supplied to the bottoming plant of the efficiency $\eta_c = 0.182$. Power output in the topping plant ($\eta_g = 0.328$) plus bottoming plant power output ($0.15 * 0.182 = 0.028$) gives a total power output per unit energy of fuel supplied $\eta = 0.328 + 0.028 = 0.356$.

On the other hand, should the gas turbine plant be able to operate without cooling, the work output (W_0) would be greater by ΔW_g then that from the topping plant (W_g) and the efficiency would be $\eta_0 = W_0/F = (W_g + \Delta W_g)/F = \eta_g + \Delta\eta_g$. This is the basic case which may be used as a comparison, in particular, to enable the decrease in efficiency due to cooling to be determined:

$$\eta_0 - \eta = \eta_0 - \eta_g - \chi_c \eta_c = \Delta\eta_g - \chi_c \eta_c . \qquad (2)$$

At the same fuel heat input of 470.6 MW the following values would be obtained: $W_0 = 178.8$ MW and $\eta_0 = 0.38$. The decrease in the work output from the topping plant as a fraction of the fuel heat supplied would be $\Delta\eta_g = (178.8 - 154.2)/470.6 = 0.052$. Part of this lost work would be restored by the bottoming plant ($W_c/F = \chi_c \eta_c = 0.028$), thus the total decrease in efficiency of the whole plant would be given by $\Delta\eta_g - \chi_c \eta = 0.052 - 0.028 = 0.024$.

An important effect that must not be neglected is the reduction in the work potential of the discharge exhaust gases. This would be reflected directly in the loss of work output from the bottoming plant in which heat is supplied from the gas turbine exhaust. The drop in the turbine exhaust temperature due to cooling is indirect evidence that the work potential is decreased. For the considered example, the reduction in this temperature is about 90°C.

3 Applying Exergy Principles

A simplified diagram of the plant shown in Fig. 2 is given in Fig. 3. Symbols are defined in the figure. After expansion of each working fluid in its own turbine, they are mixed within control volume Z. The creation of entropy of mixing is the same as in the real turbine.

For each control volume the exergy balance may be written. For control volume X the exergy balance is

$$M_a e_{a\ \mathrm{in}} + M_f e_{f\ \mathrm{in}} - W_g - M_g e_{g\ \mathrm{out}} - E_{Qg} - E_{Dg} = 0, \qquad (3)$$

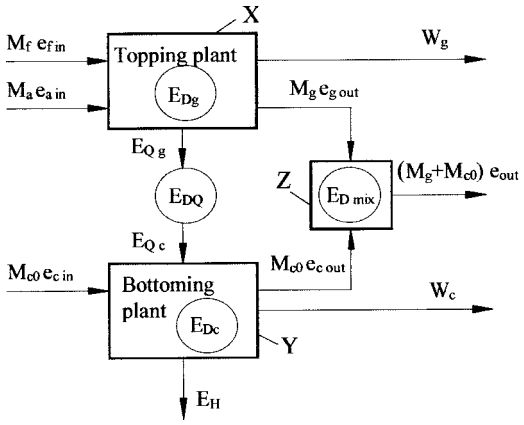


Fig. 3 Exergy balance

where E_{Qg} denotes the rate of exergy transfer accompanying the rate of heat transfer (Q_c) and represents the exergy waste due to heat rejection for the topping plant (for the gas system). In terms of available energy E_{Qg} represents the work potential of the heat rejected. The term E_{Dg} accounts for the destruction of exergy due to irreversibilities within control volume X (the combustion irreversibility; frictional pressure drops in the combustion chamber and ducting; and frictional effects in flow through the turbine and the compressor).

For control volume Y, the exergy balance may be written as

$$M_{c0} e_{c\ \mathrm{in}} + E_{Qc} - E_H - W_c - M_{c0} e_{c\ \mathrm{out}} - E_{Dc} = 0, \qquad (4)$$

where E_{Qc} represents the rate of exergy transfer with heat supplied (Q_c) for the bottoming plant (for the coolant system); the term E_{Dc} accounts for the destruction of exergy due to irreversibilities within control volume Y, the term E_H represents the rate of exergy transferred with heat extracted in an external cooler, if it takes place.

There are the following main sources of irreversibility within control volume Y: frictional effects in flows through compressors and the turbine; frictional pressure drops in air pipes between compressors and the turbine; and frictional pressure drop in cooling passages (coils in the hypothetical plant are the counterpart of cooling passages in blades and vanes of a real turbine).

As pointed out above, for the topping plant the heat extracted (Q_c) is the waste heat and the exergy (E_{Qg}) transferred with this heat represents a waste of exergy. The bottoming plant is supplied with exergy input (E_{Qc}) derived from the topping plant. The work output from the bottoming plant (W_c) is exergy in transit while the remainder of the supplied exergy is partly destroyed by irreversibilities (E_{Dc}) and is partly discharged to the surroundings ($M_{c0} e_{c\ \mathrm{out}}$ and E_H).

Since the transfer of heat between systems is not reversible, the term $E_{DQ} = E_{Qg} - E_{Qc}$ is the exergy destruction due to irreversibility in transfer of heat. The rate of destruction of exergy ($E_{DQ} + E_{Dc}$), may be interpreted as the loss of work potential due to cooling—''the cooling losses.'' To this value must be added the amount of the destruction of exergy due to mixing ($E_{D\ \mathrm{mix}}$), which inevitably occurs if two different gaseous components are mixing. Finally, the rate of the destruction of exergy due to cooling is

$$E_{D\ \mathrm{cool}} = E_{DQ} + E_{Dc} + E_{D\ \mathrm{mix}} . \qquad (5)$$

The ratio of the destruction of exergy due to cooling ($E_{D\ \mathrm{cool}}$) to the supplied exergy of fuel is a measure of inefficiency of cooling technology.

The presented equations based on the use of straightforward thermodynamics are intended for performance evaluation of the hypothetical gas turbine plant which may be regarded as a model

for an actual plant. But the science of thermodynamics is not sufficient to predict performance of a turbine with a finite number of stages at large velocities.

Heat transfer process between the gas and the coolant flows is more complicated in such turbines. Two processes, heat transfer and mixing, go on simultaneously at large velocities. As the result of mixing the gas and the coolant flows exchange momentum, the coolant accelerating to the main flow velocity. The flux of the gas momentum drops and the flux of the coolant momentum increases. The dissipation of the kinetic energy therewith occurs. The terms accounting for the change in momentum and for the destruction of exergy due to kinetic energy dissipation may be added to Eqs. (3) and (4).

4 Exergy Flow Diagram

The exergy flow diagram may be used to illustrate the thermodynamic features of an air-cooled gas turbine plant and the discussed method. As an example, consider the gas turbine plant operating in the presence of an environment at the pressure of 0.1013 MPa and the temperature of 15°C. The relative humidity of ambient air is 60 percent. The condition for the inflow air at the compressor is taken as ambient air condition. The turbine outlet stagnation pressure is assumed to be 0.1056 MPa (the diffuser is outside control surface). The mass flow rate of the air entering the compressor is 500 kg/s and the compressor pressure ratio is 15. Air is extracted from the compressor at three points and from the compressor discharge to cool turbine balding.

Methane of chemical exergy 52172 kJ/kg is supplied to the combustion chamber of 99.5 percent efficiency at a pressure of 2.28 MPa. The temperature at turbine inlet is 1260°C. The first two of four stages in the turbine are cooled. Cooling air for the first stage vanes is supplied from the compressor discharge and for the other rows from intermediate compressor bleed points. The bled air is cooled to 200°C by being passed through external coolers before being supplied to cooling channels. Maximum material temperature of vanes and blades is 800°C.

Table 1 gives details of data obtained from step by step calculation of the gas turbine. The mean section performance of the stages was determined using experimental turbine loss data, the gas-dynamic, thermodynamic, and heat transfer equations. To determine the heat flow into the each blade row, and the required cooling mass flow rate, the empirical Nusselt–Reynolds–Prandtl number relations given by Bodunov and Lokay [1] were employed.

The results presented were obtained for convective vane and blade cooling technology. It is assumed that the heat transferred by convection is removed both from the nozzle rows at constant stagnation pressure before expansion, and from the rotor rows at constant relative pressure before expansion. To simplify the analysis, it is also assumed that specific ''heat transferred from the gas to the coolant by mixing'' is equal to the stagnation enthalpy drop in the gas flow during the mixing process. It is worth emphasizing that this stagnation enthalpy drop does not correspond directly to the transferred heat, since there is, in addition, the change in the flux of momentum, that is, ''work in transit'' from the gas to the coolant flow.

Table 1 illustrates working fluid conditions, work and destruction of exergy in the air-cooled gas turbine. The total amount of exergy destroyed is $10.9+8.4+3.1+1.2=23.6$ MW. The dominant factor is the irreversibility due to profile and secondary losses in flow through turbine cascades ($10.9+1.2=12.1$ MW or 51 percent of the total amount). The irreversibility involved in transfer of heat between the gas and coolant flows is the second largest factor: $8.4*100/23.6=36$ percent of the total destruction of exergy. Destruction of exergy due to friction in cooling passages is about 13 percent of the total one. (The cooling air flows are assumed to enter cooling passages at the local pressures 0.1 MPa higher than those of the working fluid at inlet to the respective cascades.)

The rate of work of about 321.0 MW is obtained from the gas flow with the rate of heat removed of 72.7 MW. The coolant flow is supplied with this heat and produces 42.3 MW in the turbine (about 11.6 percent of the total work developed).

The locations and magnitudes of destructions of exergy for the gas turbine plant are given in Table 2. The values in the second column are given as percentages of the exergy of the fuel supplied. In the gas system, the largest irreversibility occurs in combustion. The corresponding amount of exergy destroyed is about 29 percent of the fuel exergy supplied or about 86 percent of the total value within the gas system. The most exergy destructions in a combustion chamber can be reduced by raising the turbine inlet temperature.

The 1.7 percent of fuel exergy supplied is destroyed between the gas and coolant systems, and about the same value (1.7 percent) is destroyed within the coolant system. The former is due to heat transfer over finite temperature difference, which takes place between the gas and coolant streams, and may be reduced by

Table 1 Working fluid conditions, work, destruction of exergy in cooled gas turbine

Stages, turbine	1	2	3	4	Turbine
Mass flow rate at inlet, kg./s	430.3	487.2	504.9	508.4	430.3
Stagnation conditions at inlet:					
pressure, MPa	1.47	0.76	0.38	0.19	1.47
temperature, °C	1260	987	798	642	1260
Stagnation conditions at outlet:					
pressure, MPa	0.76	0.38	0.19	0.11	0.11
temperature, °C	987	798	642	530	530
Total work output, MW	101.4	102.6	93.3	66.0	363.3
Gas system					
Flow exergy at inlet, MW	504.5	360.3	258.7	175.3	504.5
Flow exergy at outlet, MW	360.3	258.7	175.3	116.8	116.8
Heat removed, MW	56.8	12.7	2.4	0.8	72.7
Fraction of heat removed through wall	0.10	0.13	0.04	0.0	0.10
Transfer of exergy with heat transfer, MW	43.4	9.8	1.9	0.7	55.8
Destruction of exergy due to irreversibility in flow, MW[1]	3.7	2.9	2.3	2.0	10.9
Work, MW	97.1	88.9	79.2	55.8	321.0
Destruction of exergy due to irreversibility in transfer of heat, MW	6.4	1.7	0.2	0.1	8.4
Coolant system					
Flow exergy at inlet, MW[2]	13.7	47.9	41.8	28.8	17.9
Flow exergy at outlet, MW	44.3	41.3	28.7	18.7	18.7
Transfer of exergy with heat transfer, MW	37.0	8.1	1.7	0.6	47.4
Destruction of exergy due to friction in cooling passages, MW	2.0	0.7	0.3	0.1	3.1
Destruction of exergy due to irreversibility in flow, MW	0.1	0.3	0.4	0.4	1.2
Work, MW	4.3	13.7	14.1	10.2	42.3

[1]These values include destructions of exergy due to dissipation of kinetic energy during the mixing process. The overall rate of destructions is 1.2 MW.

[2] These values are equal to the sum of exergy transfer rate with coolant mass flow entering the stage as working fluid and that entering cooling passages.

Table 2 Destruction of exergy in cooled gas turbine plant

	in MW	as %
Gas system		
Compressor	11.1	2.3
Combustion chamber and fuel supply	145.5	29.4
Turbine (Table 1)	10.9	2.2
Totals	167.5	33.9
Between systems (Table 1)	8.4	1.7
Coolant system		
Compressors	2.0	0.4
Air lines between compressors and turbine	1.8	0.4
Turbine (Table 1)	4.3	0.9
Totals	8.1	1.7

Table 3 Exergy balances

	in MW	as %
Gas system		
Exergy of fuel	494.3	100.0
Flow exergy at compressor inlet	0.0	0.0
Transfer of exergy with heat transfer,	55.8	11.3
Destruction of exergy within system	167.5	33.9
Flow exergy at turbine outlet	116.8	23.6
Work output	154.2	31.2
Destruction of exergy between systems	8.4	1.7
Coolant system		
Flow exergy at compressors inlet	0.0	0.0
Transfer of exergy with heat transfer	47.4	9.6
Destruction of exergy within system	8.1	1.7
Loss of exergy due to heat rejection in coolers	7.1	1.4
Flow exergy at turbine outlet	18.7	3.8
Work output	13.2	2.7
Destruction of exergy due to mixing	9.4	1.9
Exergy of exhaust gases after mixing	126.1	25.5

increasing blade heat exchange effectiveness. In the coolant system, pressure losses can contribute significantly to the overall rate of the destruction of exergy. This value is about of 24 percent (0.4*100 percent/1.7=24 percent) of the total one and may be minimized by selection, of optimum extraction points in the compressor. If coolant flows are cooled in external coolers, then there is loss of exergy due to heat rejection. In the plant being considered, this loss is about 1.4 percent of the supplied fuel exergy.

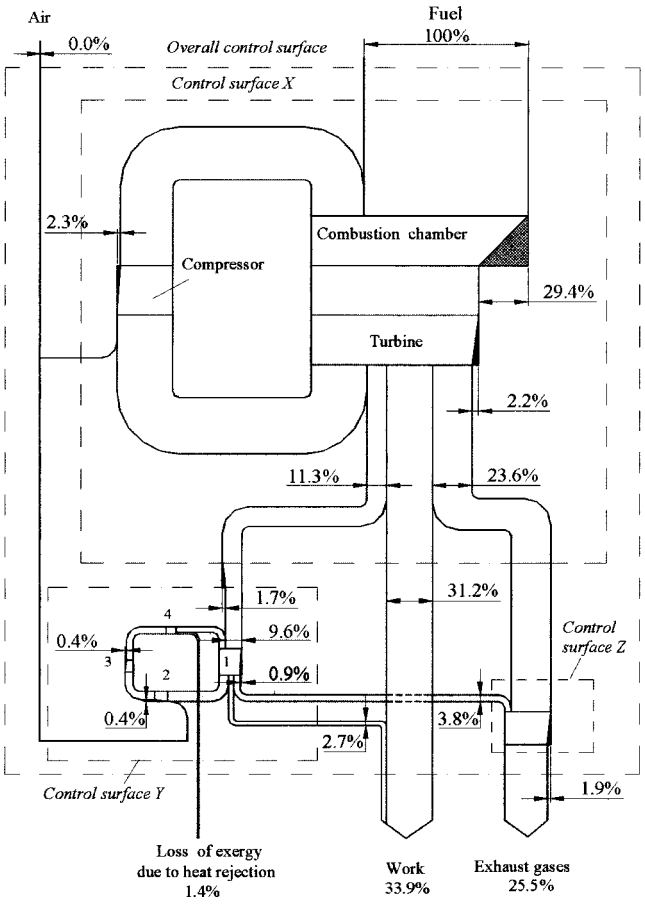


Fig. 4 Exergy flow diagram. 1, turbine. 2, compressors. 3, air lines between compressors and turbine. 4, coolers

Table 3 shows exergy balances for each of system and for the whole plant. Figure 4 shows the exergy flow diagram for such plant. The exergy of the fuel supplied is taken as being 100 percent. The specific exergy of ambient air is assigned to zero. The destruction of exergy due to fuel throttling is added to that in the combustion chamber.

The exergy balance indicates that 31.2 percent of the exergy of fuel is converted into work by the gas system, about 33.9 percent is destroyed within the system, and 11.3 percent is transferred to the coolant system. The rest, 23.6 percent, is discharged with exhaust gas stream.

Since 1.7 percent of supplied fuel exergy is destroyed between systems, 9.6 percent enters the coolant system, of which 2.7 percent is obtained as work; 5.2 percent (3.8 percent+1.4 percent =5.2 percent) is discharged with exhaust stream and is transferred with heat in coolers. The balance, 1.7 percent, is destroyed within the system.

About 1.9 percent of supplied fuel exergy is destroyed due to mixing of exhaust streams. This value was calculated from the exergy balance for the mixing process, as shown in Fig. 3. The chemical exergy of fuel and the thermal exergy of gases were calculated in the way described by Shargut et al. [2].

Thus the necessity to cool turbine blading in considered plant has led to the destruction of exergy in amount of 1.7 percent +1.7 percent+1.9 percent=5.3 percent of the fuel exergy supplied and to the loss of exergy due to precooling in the amount of 1.4 percent.

5 Conclusions

In the thermodynamic sense, cooling losses are caused by irreversibilities which result from the modifications to the basic cycle (circuit) to a turbine cool.

An air-cooled gas turbine plant may be treated as a plant in which a combination of working fluids is used. A circuit on which such a plant operates may be treated as a compound circuit comprising the open circuit with internal combustion and heat rejection during expansion, which is superposed directly on the other circuit supplied with heat from the former. The latter uses a coolant as working fluid.

Cooling losses derive both from the irreversible heat transfer between circuits and from irreversible internal processes through the coolant goes.

The type of analysis given in the paper for the simple air-cooled gas turbine plant may be extend to any cooled gas turbine plant that derives its energy from the combustion of organic fuel, whether the system involves regeneration, reheat, or intercooling. To this end, it is necessary to consider replacement of the existing plant by the modified plant consisting of two plants. The first plant produces work from the chemical energy of the fuel and uses air and products of combustion as working fluids. The heat extracted from products of combustion during the cooled expansion process in one or more turbines is used as heat supply to the second plant producing work and using the coolant as working fluid. The overall efficiency of the modified plant will be the same as that of a considered plant.

The presented type of analysis may be interpreted by employing the concept of availability, allowing all irreversibilities to be identified and quantified. No approach other than the exergy method does appear to offer such possibilities for analyzing cooled gas turbine plant.

Nomenclature

Q = heat
W = work
M = mass flow rate
η = efficiency

E, e = flow exergy, specific flow exergy
E_D = destruction of exergy

Subscripts

a = air
f = fuel
g = products of combustion (gas); referring to the topping plant
c = coolant; referring to the bottoming plant

References

[1] Bodunov, M. N., and Lokay, B. I., 1971, ''The External Heat Transfer Coefficient,'' High-Temperature Cooled Gas Turbines, Machinostroenye, Moscow, pp. 23–44.
[2] Szargut, J., Morris D. R., and Steward, F. R., 1988, *Exergy Analysis of Thermal, Chemical, and Metallurgical Processes*, Hemisphere, New York.
[3] Kotas, T. J., 1985, *The Exergy Method of Thermal Plant Analysis*, Butterworths, London.
[4] Moran, M. J., and Sciubba, E., 1994, ''Exergy Analysis: Principles and Practice,'' ASME J. Eng. Gas Turbines Power, **116**, pp. 285–290.

Experimental Assessment of Fiber-Reinforced Ceramics for Combustor Walls

D. Filsinger
S. Münz
A. Schulz
S. Wittig

Lehrstuhl und Institut für Thermische Strömungsmaschinen, Universität Karlsruhe (T.H.), Kaiserstraße 12, 76128 Karlsruhe, Germany

G. Andrees
Motoren- und Turbinen Union München GmbH, Dachauerstraße 665, 80995 München, Germany

Experimental and theoretical work concerning the application of ceramic components in small high-temperature gas turbines has been performed for several years. The significance of some nonoxide ceramic materials for gas turbines in particular is based on their excellent high-temperature properties. The application of ceramic materials allows an increase of the turbine inlet temperature resulting in higher efficiencies and a reduction of pollution emissions. The inherent brittleness of monolithic ceramic materials can be virtually reduced by reinforcement with ceramic fibers leading to a quasiductile behavior. Unfortunately, some problems arise due to oxidation of these composite materials in the presence of hot gas flow containing oxygen. At the Motoren und Turbinen Union, München GmbH, comprehensive investigations including strength, oxidation, and thermal shock tests of several materials that seemed to be appropriate for combustor liner applications were undertaken. As a result, C/C, SiC/SiC, and two C/SiC composites coated with SiC, as oxidation protection, were chosen for examination in a gas turbine combustion chamber. To prove the suitability of these materials under real engine conditions, the fiber-reinforced flame tubes were installed in a small gas turbine operating under varying conditions. The loading of the flame tubes was characterized by wall temperature measurements. The materials showed different oxidation behavior when exposed to the hot gas flow. Inspection of the C/SiC composites revealed debonding of the coatings. The C/C and SiC/SiC materials withstood the tests with a maximum cumulated test duration of 90 h without damage. [DOI: 10.1115/1.1364523]

Introduction

Gas turbine development today concentrates on higher thermal efficiencies and lower pollution emissions. One possible approach is an increase of turbine inlet temperature. Therefore, either an effective cooling of the components forming the hot section or high-temperature resistant materials are required. Carbon (C) or silicon carbide (SiC) based materials with their outstanding high-temperature properties offer a great potential for such applications [1,2].

However, the advantageous qualities of monolithic materials in structural applications, such as their high Young's modulus, high thermal conductivity, and excellent oxidation resistance, are partly offset by their relatively low fracture toughness. A wide range of additives and particulate reinforcements have been tried as toughening agents. Reinforcement with SiC or C continuous fibers could yield the highest fracture toughness. These materials show an increase in fracture toughness by one order of magnitude and an increase in fracture energy by two orders of magnitude [3]. Since they have to operate under hostile stress-temperature environment conditions, the choice of toughening phases is restricted by high-temperature chemical incompatibility and by thermal expansion coefficient mismatch with the matrix.

In addition, the desired applications involve extended periods of operating in oxidizing atmospheres which require a proven oxidation resistance of the composites [4]. Therefore, at the Institut für Thermische Strömungsmaschinen (ITS), in cooperation with the Motoren und Turbinen Union München GmbH, different SiC and C-based composite materials were tested for their oxidation behavior. In preliminary investigations, the materials exposed to high temperatures in an atmosphere containing oxygen were judged by mass reduction and strength tests. The most promising materials were selected for manufacturing flame tubes to test their durability under real engine conditions. The new flame tubes were tested in the ceramic combustion chamber of the Klöckner Humboldt Deutz T216-type gas turbine that was designed at the ITS [5].

Materials Selection

Three different composite materials were chosen for the tests under real engine conditions. These materials were selected out of seven SiC or C composites and four fiber-reinforced glasses. They offered the most promising high-temperature properties which are necessary for combustor walls. For example, the glass materials suffered from a drastic strength loss for temperatures exceeding 1000°C [6]. For reasons of brevity, in the following section only the preliminary test results of the materials selected are shown (see Table 1).

Under oxidizing atmospheres, fiber degradation was expected in the two-directional composites reinforced with silicon carbide fibers. In the case of the materials with carbon fibers, rapid oxidation was expected. The mass reduction of the composites reflects the structural changes of the specimens. The qualification was done by strength measurements with respect to the mass loss. The experiments were done in a muffle furnace at elevated temperatures. Figure 1 shows the mass reduction of the carbon reinforced materials versus time for two different temperatures, and in Fig. 2 the strength dependence on mass reduction is demonstrated. Hence a life expectancy can be determined if the lowest allowable strength is prescribed.

Assuming that a minimum tensile strength of 150 MPa is sufficient for reliable operation of the component, the durability in a 1000°C atmosphere would be only 8 min for the C/C material and only 6 min for the C/SiC material. This clearly shows the need for effective external oxidation protection.

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Orlando, FL, June 2–5, 1997; ASME Paper 97-GT-154. Manuscript received by IGTI Dec. 1996; final revision received by the ASME Headquarters Mar. 1997. Associate Editor: H. A. Kidd.

Table 1 Selected fiber reinforced materials

Composite	Interface coating	Fabrice (layup)	Manufacturer
C/C	-	satin weave (2D-0/90°)	Schunk
C/SiC	PyC	UD laminate (2D-0/90°)	Dornier
SiC/SiC	PyC	plain weave (2D-0/90°)	MAN Technologie AG

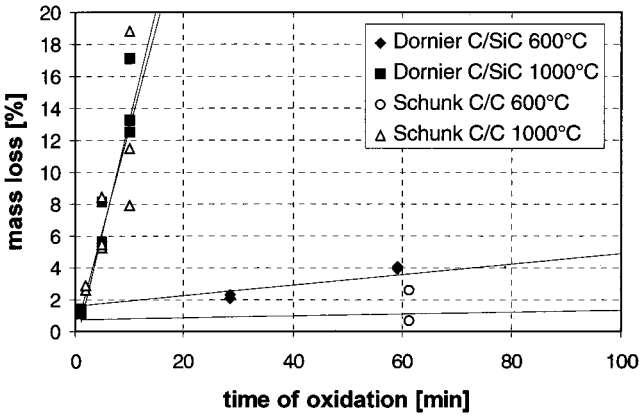


Fig. 1 Mass loss versus oxidation time for composite materials

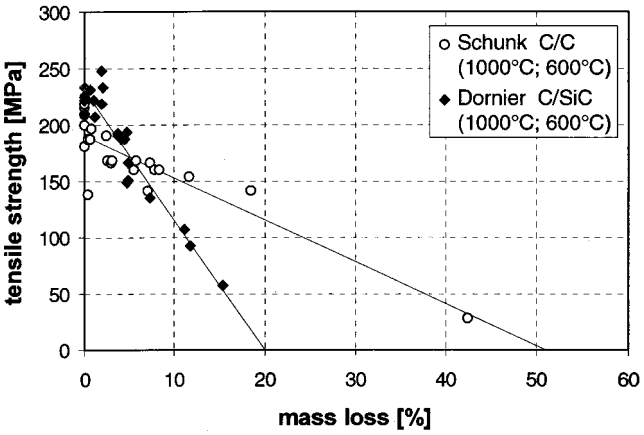


Fig. 2 Strength dependence on mass loss

Table 2 Oxidation behavior of coated composites

	Composite	Mass loss [%]	Tensile strength [MPa]
as delivered	C/C	-	90-93
1230°C/ 1000 h	C/C	8.1-8.6	85-99
1430°C/ 100 h	C/C	7.8-8.4	101-130
as delivered	C/SiC	-	198-252
1230°C/ 1000 h	C/SiC	7.5-25.5	destroyed
1430°C/ 100 h	C/SiC	8.5-18.9	77-96
as delivered	SiC/SiC	-	284.7
1230°C/ 1000 h	SiC/SiC	-	117-130
1300°C/ 100 h	SiC/SiC	0.22-0.37	140-156

Table 3 Flame tube materials

Composite	Oxidation protection	Manufacturer
C/C	CVD-SiC	Schunk
C/SiC	CVD-SiC	Dornier
C/SiC	dross SiC	Dornier
SiC/SiC	CVD-SiC	MAN Technologie AG

Since coatings can influence the mechanical properties of the base material, strength tests were conducted for coated specimens. It was demonstrated that SiC-coated C/C and C/SiC showed reduced tensile strengths depending on the coating process. This is a result of the brittle surface and an increase in the cross sectional area of the specimen caused by the coating that has poor mechanical properties. However, the protected materials showed a significant reduction in mass loss while maintaining comparatively high tensile strengths, when exposed to the hot gas atmosphere (see Table 2).

Since the carbon-reinforced composites did not show sufficient oxidation resistance, a SiC/SiC material with a SiC and an additional glaze-based silicate (SiO_2) coating was added to the test program. The oxidation test results are also listed in Table 2.

Comparing the C- and SiC-reinforced materials, the superiority of SiC/SiC can be clearly seen. The reason for this superiority is that the degradation of the SiC fibers seems to be not as critical as the oxidation of the carbon fibers.

Out of these materials four flame tubes were manufactured for tests under engine conditions (see Table 3). The fiber architecture and the interface coating correspond to the tested specimens. The carbon-reinforced composites were coated with SiC either by chemical vapor deposition (CVD) or a dross process. The SiC/SiC material was protected by a CVD-SiC layer.

It has to be mentioned that the protected C/C composite for the engine tests does not conform with the material listed in Table 2, since this material suffered from poor mechanical properties. According to Schunk Company the flame tube material exhibits a much better strength due to a modified CVD process for the coating of the base material.

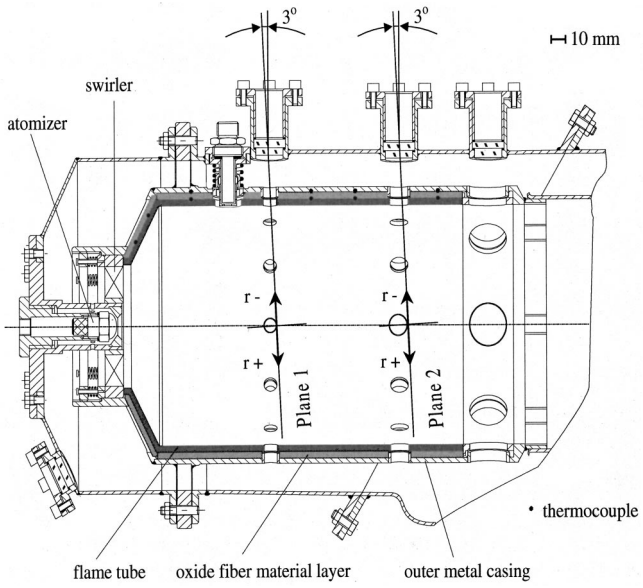


Fig. 3 Flame tube construction

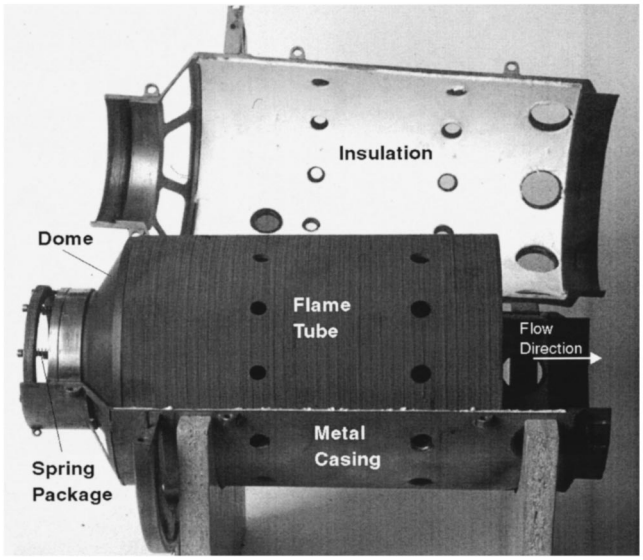


Fig. 4 Ceramic flame tube

Combustion Chamber

A cross-sectional view of the ceramic-can-type combustor is given in Fig. 3. The combustor walls consist of three layers. The inner layer is a hot gas resistant composite, the middle layer is a flexible oxide fiber material, and the outer layer is a metal casing. The flexible insulation in the radial direction, and the spring support of the swirler in the axial direction allows for an almost unhindered thermal expansion of the ceramic flame tube. This reduces the critical tensile stresses to a minimum, especially when considering the thermal expansion mismatch of the metal containment and the composite liner wall. The insulation keeps the fiber reinforced material at a more homogeneous temperature distribution and lowers the thermal stress level. The temperature loading of the combustor liner is measured by thermocouples along the outer ceramic wall. Flanges on the outer side of the combustion chamber give access to the flame tube for measuring the radial temperature distribution with transversable PtRh-Pt thermocouples (not shown in Fig. 3). The measuring planes are declined by 3 deg to detect the streaks of the cold dilution air.

This newly designed ceramic combustion chamber was integrated into a Klöckner Humboldt Deutz T216-type gas turbine.

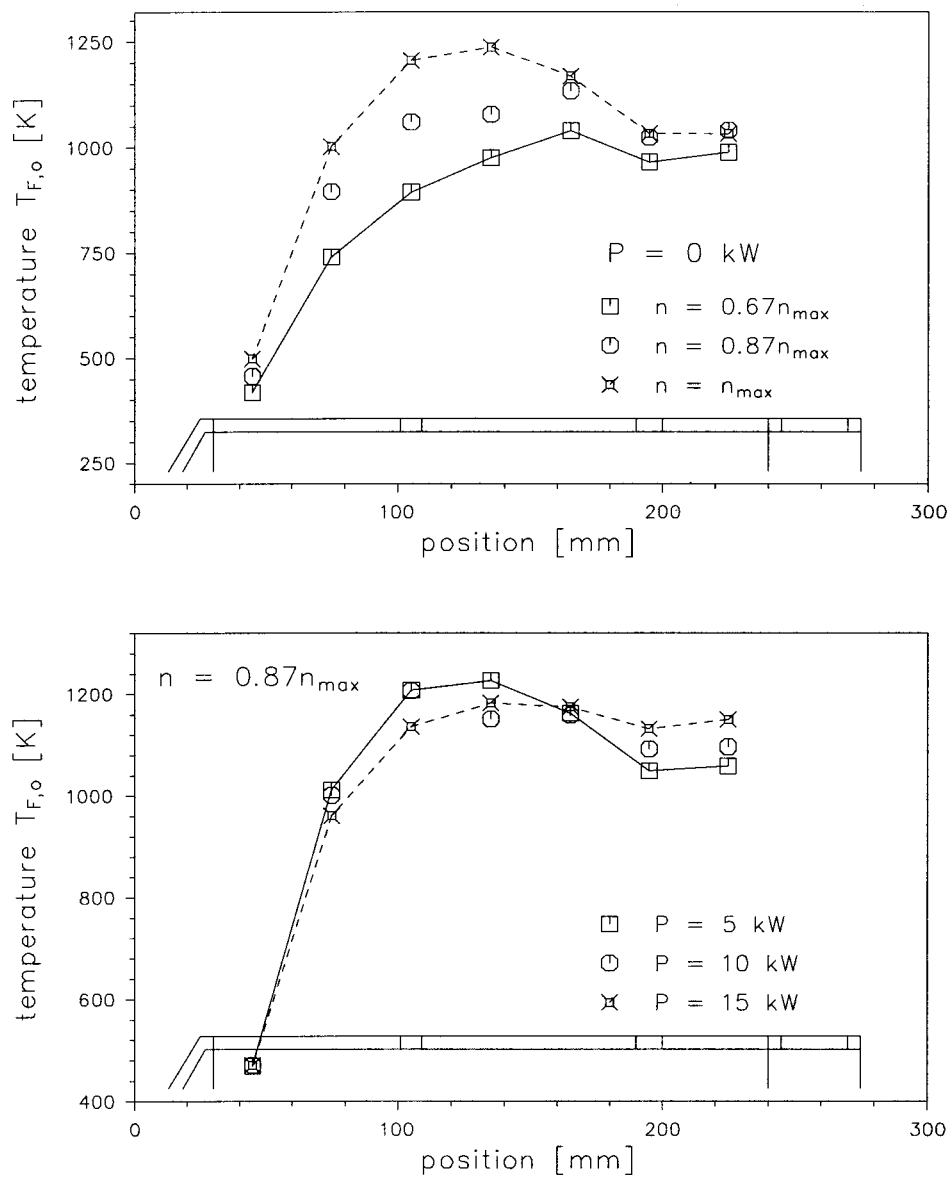


Fig. 5 Temperature distributions on the outer surface of the flame tube

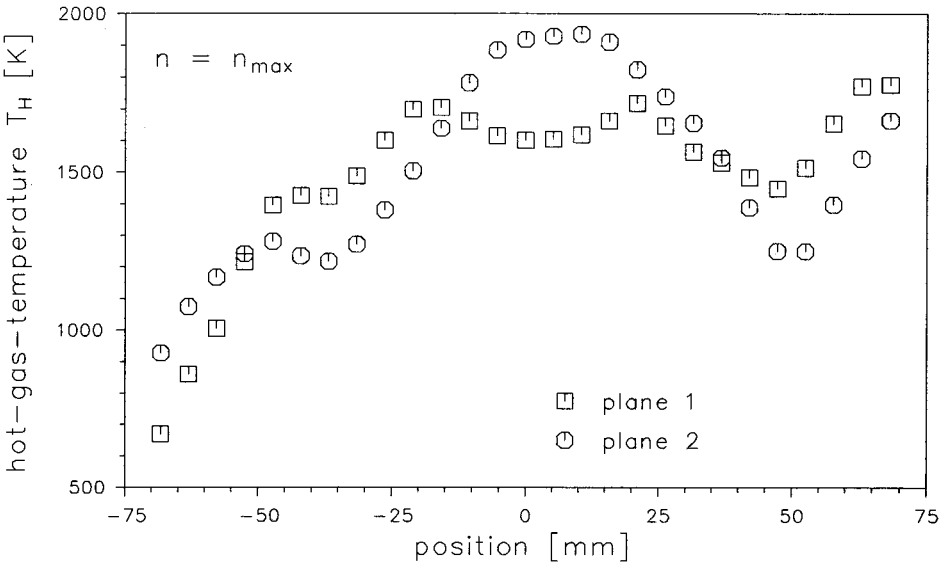


Fig. 6 Radial hot-gas temperature distribution in the flame tube

The output power of the engine is 74 kW at 50,000 rpm. The nominal pressure ratio is 2.8 and the air mass flow is 0.9 kg/s at a turbine inlet temperature of 810°C.

The dome of the combustor consists of sintered silicon carbide. Since large holes for the dilution air are necessary at the end of the flame tube, this part is separated from the ceramic flame tube and made of a nickel alloy. The flame tube itself has a length of 210 mm and an inner diameter of 144 mm. The thickness of the ceramic wall is 3 mm. The fiber reinforced flame tubes were made by the manufacturers indicated in Table 3. The pipe specimens were installed into the combustion chamber and exposed to the hot gas flow under real engine conditions. Figure 4 shows a photograph of the flame tube assembled with the SiC/SiC tube from MAN Technology.

Experimental Results

During machine operation, all characteristic parameters such as pressures, temperatures, speed, power, etc., were recorded. The wall temperatures were used to determine the thermal loading of the composite materials. Figure 5 displays typical axial temperature distributions of the ceramic wall. The lines refer to different operating conditions. As expected, the thermal loading increases with higher rotational speeds. Increasing the output power also causes higher-temperature loading, especially near the exit of the flame tube. This is due to the elongation of the flame under these conditions.

The gas turbine was run under various operating points. The highest outer wall temperature observed was about 1050°C and was detected between the dilution holes in the middle of the flame tubes at nominal speed. The temperature distributions were similar for all flame tubes tested.

The accumulated testing time for the carbon reinforced composites was limited to 10 h. At first, operating points referring to lower thermal loads were tested. With increasing test duration the thermal load was intensified. Between the single tests the flame tubes were inspected visually and morphological changes were documented by macro-photography. Hot-gas profiles were measured for all operating conditions. An example of the radial temperature distribution is plotted in Fig. 6. This graph shows that for nominal speed the flame establishes right between the measurement planes. The influence of the injected air can be seen at relatively cold zones near the wall.

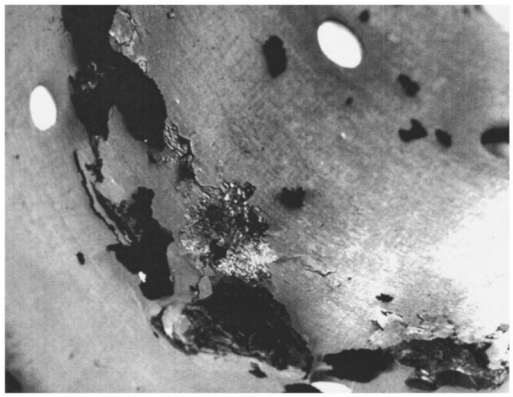


Fig. 7 C/SiC flame tube with dross SiC coat after 10 h of engine operation

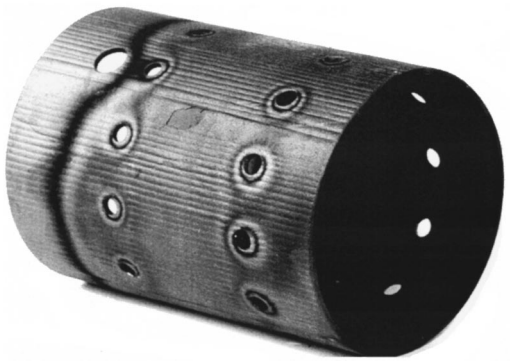


Fig. 8 C/SiC flame tube with CVD-SiC coat after 10 h of engine operation

All four flame tubes withstood the thermal load under an oxidizing atmosphere without severe damage. The Dornier flame tube, coated with the dross SiC layer, showed debonding on the inner surface at the highest load section. Since preliminary studies revealed that this material does not have a long life expectancy without oxidation protection, this coating and composite combination cannot guarantee reliable long-term machine operation (see Fig. 7).

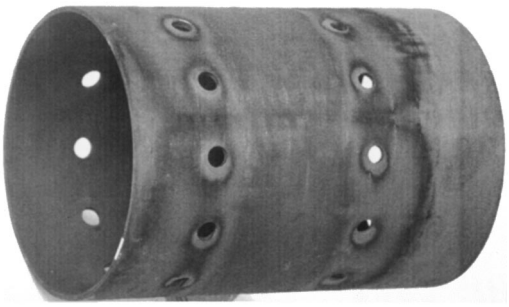


Fig. 9 C/C flame tube with CVD-SiC coat after 10 h of engine operation

Fig. 10 Inner side of the SiC/SiC flame tube after 30 h of engine operation

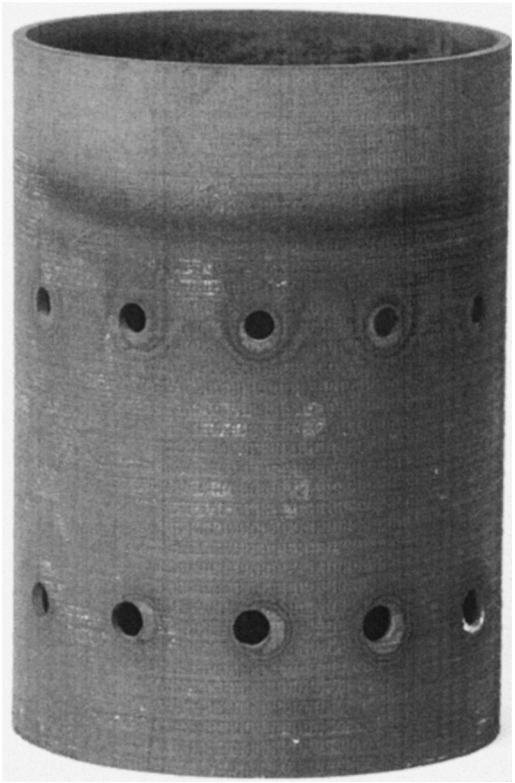


Fig. 11 SiC/SiC flame tube after 50 h of engine operation

The second Dornier flame tube coated with CVD-SiC showed no damage on the inner side. However, on the outer side a small part of the coat chipped off. This is believed to be a result of a mismatch between the thermal expansion coefficients. The tube was still in good condition, after operating for 10 h, and did not exhibit further damage. The thermally high-loaded regions showed an intensive discoloring. This phenomenon was detected on all flame tubes except the one coated with dross SiC. The discolorations are due to establishing a SiO_2 layer which results from the oxidation process and helps to suppress the subsequent oxidation. The thin amorphous glass layers reflect the different wavelengths of the incoming light depending on their thickness. As a result, the surface appears with rainbowlike colors (Fig. 8).

The C/C flame tube that was manufactured by Schunk withstood testing with no structural defects. Only the surface color had changed after 10 h of operation (Fig. 9).

The test program was extended for the promising SiC/SiC composite. Besides the alteration in surface color, no significant change was detected after 10 h. The only significant difference to the ''as-delivered'' state was a small area of debonding at the front face of the pipe. This was due to chemical reactions between the Ni-based metallic ring element and the ceramic composite material. The manufacturer, MAN Technology AG, observed similar reactions in other applications. Therefore, in the following tests, the SiC composite was separated from the metal by a thin interface layer which consisted of a fiber oxide material.

The SiC/SiC combustion chamber, in operation now for more than 90 h, exhibited no damage. It also had to withstand numerous start and shut down processes which caused critical loads because of the high-temperature gradients. Downstream of the dilution holes, the surface was colored brown. This was probably a result of deposits from the incoming air. In Fig. 10 the area downstream of a hole in the flame tube is shown. The position of soot deposits also demonstrate the tangential velocity component of the swirling air flow in the combustion chamber.

The tests under real engine conditions seem to confirm the superior properties of this silicon carbide reinforced composite. The material withstands the oxidizing atmosphere with no structural change, as can be seen in Fig. 11.

Conclusions

Four flame tubes made of composite materials with a variety of coatings for oxidation protection were tested under real turbine conditions. Preliminary studies by the Motoren und Turbinen Union, München GmbH, helped in identifying the most promising composites. The flame tubes were installed into the ceramic combustion chamber developed at the Institut für Thermische Strömungsmaschinen.

First examinations of the uncoated composites showed that oxidation of the C fibers was a very rapid process, and it was concluded that oxidation protection is necessary [7]. Tests with coated specimens in temperature regimes between 1200 and 1400°C revealed that the degradation of the SiC fibers is less critical. As a result, the SiC/SiC material showed the best oxidation resistance.

After operating 10 h in the Klöckner Humboldt Deutz T216-type gas turbine, both C/SiC flame tubes showed debonding of the different coatings. This debonding was probably due to the differing thermal expansion coefficients. The C/C material coated with CVD-SiC withstood the 10 h test program with no damage. For the SiC-reinforced SiC composite, the test program was extended since this material promised good oxidation resistance. The maximum outer wall temperatures during the tests were 1050°C. The component was under operation for 90 h and visual inspection showed no morphological defects.

Tests for the C/C- and SiC/SiC-flame tubes are still continued. The project is planned to be completed with destructive tests of the tubes to obtain any structural changes in the materials.

Acknowledgments

The authors are grateful to the ''Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie'' which supported the program in the scope of the ''Arbeitsgemeinschaft Hochtemperaturgasturbine'' (AG-Turbo) under Contract No. 0326760A.

References

[1] Strife, J. R., and Sheehan, J. E., 1988, ''Ceramic Coatings for Carbon-Carbon Composites,'' Ceram. Bull., **67**, pp. 369–374.
[2] Lamouroux, F., and Camus, G., 1994, ''Oxidation Effects on the Mechanical Properties of 2D Woven C/SiC Composites,'' J. Eur. Ceram. Soc., **14**, pp. 177–188.
[3] Helmer, T., Peterlik, H., and Kromp, K., 1995, ''Coating of Car-bon Fibers-The Strength of the Fibers,'' J. Am. Ceram. Soc., **78**, pp. 133–136.
[4] Kirk, G. E., 1989, ''Composite Materials for Future Aeroengines,'' ASME-Paper 89-GT-313.
[5] Münz, S., Schulz, A., and Wittig, S., 1996, ''Evaluation of a New Design Concept for a Ceramic Flame Tube under Engine Conditions,'' ASME-Paper 96-GT-498.
[6] Andrees, G., Wittig, S., Schulz, A., Münz, S., and Filsinger, D., 1996, ''Untersuchung von faserkeramischen Werkstoffen zur Brennkammerauskleidung,'' Abschlußbericht, AG-Turbo (Turbotherm II) Vorhaben 2.4.1.4.
[7] Filsinger, D. and Münz, S., 1996, ''Untersuchung faserkeramischer Werkstoffe für Brennkammerauskleidungen unter gasturbinentypischen Bedingungen,'' Jahresbericht: Institut für Keramik im Maschinenbau, Universität Karlsruhe, pp. 147–152.

Slow Crack Growth Analysis of Advanced Structural Ceramics Under Combined Loading Conditions: Damage Assessment in Life Prediction Testing

S. R. Choi[1]
Ohio Aerospace Institute,
Cleveland, Ohio 44142
e-mail: Sung.R.Choi@grc.nasa.gov

J. P. Gyekenyesi
NASA Glenn Research Center,
Cleveland, Ohio 44135
e-mail: John.P.Gyekenyesi@grc.nasa.gov

Slow crack growth analysis was performed with three different loading histories including constant stress-rate/constant stress-rate testing (Case I loading), constant stress/constant stress-rate testing (Case II loading), and cyclic stress/constant stress-rate testing (Case III loading). Strength degradation due to slow crack growth and/or damage accumulation was determined numerically as a function of percentage of interruption time between the two loading sequences for a given loading history. The numerical solutions were examined with the experimental data determined at elevated temperatures using four different advanced ceramic materials, two silicon nitrides, one silicon carbide and one alumina for the Case I loading history, and alumina for the Case II loading history. The numerical solutions were in reasonable agreement with the experimental data, indicating that notwithstanding some degree of creep deformation presented for some test materials slow crack growth was a governing mechanism associated with failure for all the test materials. [DOI: 10.1115/1.1365160]

Introduction

Advanced ceramics are candidate materials for high-temperature structural applications in heat engines and heat recovery systems. One of the major limitations of these materials in high-temperature applications is delayed failure, where slow crack growth (also called ''fatigue'' or ''subcritical crack growth'') of inherent flaws can occur until a critical size for catastrophic failure is attained. Therefore, it is important to evaluate accurately slow crack growth (SCG) behavior with a specified loading condition so that reasonable life prediction of ceramic components is ensured.

There are several methods of determining SCG of advanced ceramics. Typically, the SCG of ceramics is determined by applying constant stress-rate (also called ''dynamic fatigue''), constant stress (also called ''static fatigue'' or ''stress rupture'') or cyclic stress (also called ''cyclic fatigue'') to ground specimens or to precracked fracture mechanics specimens in which the crack velocity measurements are made. Constant stress-rate testing determines the strength for a given applied stress; whereas, constant stress and cyclic stress testing measures time to failure for given constant stress and cyclic stresses, respectively. Of these test methods, constant stress-rate testing has been widely utilized for decades to characterize SCG behavior of ceramic materials at both ambient and elevated temperatures. The advantage of constant stress-rate testing over other methods lies in its simplicity: Strengths are measured in a routine manner at four to five applied stress rates by applying either displacement-control mode or load-control mode. The SCG parameters for life-prediction design are simply calculated from a relationship between strength and applied stress rate. Because of its advantages, constant stress-rate testing has been developed as ASTM test standards (C1368 and C1465) to determine SCG parameters of advanced ceramics at ambient temperature ([1]) and elevated temperatures ([2]).

One of the difficulties possibly encountered in elevated-temperature testing is that, depending on test conditions (test rate, time, temperature and environment) and material, the identification of a governing failure mechanism may be obscured by the presence of possible multiple mechanisms, particularly with a combination of SCG and creep ([3–6]). Thus, the determined SCG parameters cannot be solely representative of one single process, slow crack growth, but a combination of the two competing mechanisms. They may also act in series, i.e., creep followed by SCG. The underlying basis of the aforementioned SCG testing—constant stress rate, constant stress and cyclic stress testing—is the crack velocity formulation in which crack propagation typically follows a power-law relation. If the SCG mechanism is dominant for a given material temperature environmental system, then the SCG parameters obtained, in principle at least, should be in a reasonable range of accuracy, regardless of test method. Furthermore, one must be able to predict life and/or strength from any loading history that could be a combination of constant stress-rate/constant stress-rate, constant stress/constant stress-rate, or cyclic stress/constant stress-rate loading sequences. There have been some experimental attempts to evaluate the degree of crack growth or damage accumulation by determining ''fast''-fracture ''residual'' strength of silicon nitride specimens that had been subjected to and then interrupted from tensile cyclic loading at elevated temperature ([7]). However, in general, both analytical work and systematic experimental data on this subject rarely exist in the literature.

Consequently, the purpose of this paper is to better understand how damage (SCG creep or both) was accumulated with time for given loading history leading to failure of advanced structural ceramics at elevated temperatures. Numerical solutions of strength

[1]NASA Senior Resident Research Scientist, Glenn Research Center, Cleveland, OH 44135 (All correspondences to this address)

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Munich, Germany, May 8–11, 2000; Paper 00-GT-062. Manuscript received by IGTI Oct. 1999; final revision received by ASME Headquarters Oct. 2000. Associate Editor: D. Wisler.

degradation in conjunction with crack growth were obtained for each loading history with a major assumption that the governing failure mechanism was slow crack growth. Included in the test matrix were two typical methods of constant stress-rate (dynamic fatigue) and constant stress (''static fatigue'' or ''stress rupture'') testing. The SCG and related parameters were determined on the basis of these test results. Then, a combination of two different loading sequences was applied to test specimens and the corresponding strengths were measured to see how strength degradation in the form of SCG/damage accumulation took place during the combined loading sequences. The combination of loading used in this testing included slow test rate/fast test rate (which is a combination of constant stress-rate and constant stress-rate testing, called here Case I loading history) and static loading/fast test rate (a combination of static stress and constant stress-rate testing, called Case II loading). The testing was interrupted after the first loading sequence at a specified time, and then the specimens were fractured at the second loading sequence using a fast test rate of typically 33 MPa/s. The experimental results were compared with the numerical solutions.

Theoretical Background

Approach. Numerical solutions of strength, crack size, and other required variables for various loading histories, are presented in this section. The schematic loading history considered in this paper is depicted in Fig. 1. The first one, Fig. 1(a), called *Case I loading*, was a combination of two constant stress-rate testing with a fast test rate after a slow test rate. A specimen was subjected to a low stress rate. Then, the testing was interrupted at a specified time J_{int}, and resumed with a fast stress rate until the specimen fractured. The second loading history, Fig. 1(b), called *Case II loading*, was a combination of constant stress and constant stress-rate testing. The testing was started initially with a constant stress, interrupted at a specified test time J_{int} and then resumed by applying a fast stress rate until the specimen broke. The third loading history, Fig. 1(c), called *Case III loading*, which was a combination of cyclic stress and constant stress-rate testing, was simply a replacement of a static stress used in the Case II loading with cyclic stress. The ratio (φ) of interruption time to time to failure is defined as follows:

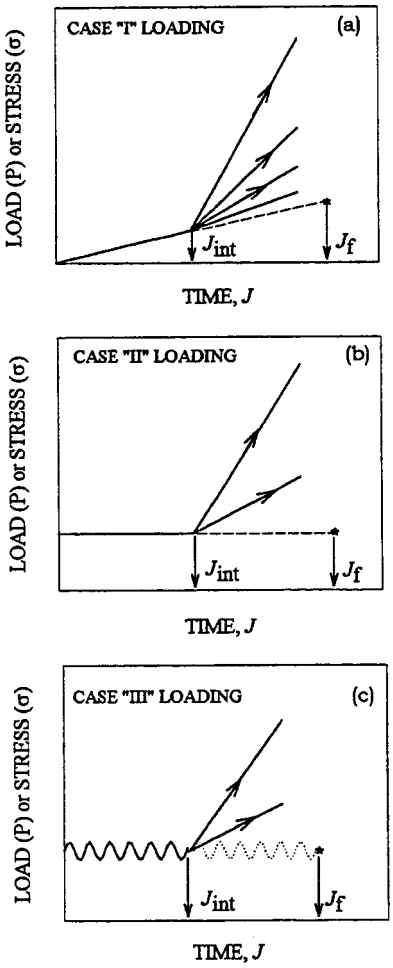


Fig. 1 Schematics of three loading histories considered: (a) Case I loading, (b) Case II loading, and (c) Case III loading

$$\varphi = \frac{J_{\text{int}}}{J_f} \tag{1}$$

where J_{int} is the interruption time and J_f is time to failure of a test specimen, subjected to only the first loading sequence (without the second loading, i.e., $\varphi=1$) which is either constant stress-rate, constant stress or cyclic stress, see Fig. 1. The case for $\varphi=0$ represents the second loading sequence with no the first loading sequence. The interruption time was chosen such that the ratio ranged typically from $\varphi=0$ to 90 percent.

In many cases, slow crack growth of advanced ceramics under mode I loading above the fatigue limit, either by stress corrosion at ambient temperature or by grain boundary sliding at elevated temperatures, can be expressed by the following empirical power-law relation ([8])

$$v = \frac{da}{dt} = A\left[\frac{K_I}{K_{IC}}\right]^n \tag{2}$$

where v, a, and t are crack velocity, crack size, and time, respectively. A and n are the material environment-dependent SCG parameters. K_I is the mode I stress intensity factor (SIF), and K_{IC} is the critical SIF or fracture toughness of the material subjected to mode I loading. The simplistic analytical solutions of strength in constant stress-rate testing and of time-to-failure in constant stress and cyclic stress testing can be approximated as follows ([9–11]):

$$\sigma_f = D_d[\dot{\sigma}]^{1/1+n} \tag{3}$$

$$t_{fs} = D_s[\sigma]^{-n} \tag{4}$$

$$t_{fc} = D_c[\sigma_{\max}]^{-n} \tag{5}$$

where σ_f is the fracture strength corresponding to the applied stress rate ($\dot{\sigma}$) in constant stress-rate testing, t_{fs} is the time to failure subjected to a constant applied stress (σ) in constant stress testing, and t_{fc} is the time to failure subjected to the maximum applied stress ($\sigma_{\max}$) in cyclic stress testing. The parameters D's can be expressed as follows ([9–11]):

$$D_d = [B(n+1)S_i^{n-2}]^{1/(n+1)}$$

$$D_s = BS_i^{n-2} \tag{6}$$

$$D_c = (BS_i^{n-2})/\left[\frac{1}{\tau}\int_0^t [f(t)]^n dt\right]$$

where S_i is the inert strength and $B=2K_{IC}/[AY^2(n-2)]$ with Y being the crack geometry factor in the relation of $K_I = Y\sigma a^{1/2}$. $f(t)$ is a periodic function in cyclic loading specified in $\sigma(t) = \sigma_{\max} f(t)$ with a range of $0 \leqslant f(t) \leqslant 1$, and τ is the period. The SCG parameters n and B (or A) can be obtained by a linear regression analysis with experimental data in conjunction with an appropriate equation, Eqs. (3), (4), or (5), depending on the type of loading.

To obtain more generalized, convenient, and accurate SCG analysis, several parameters that are commonly specified in the conventional analytical solutions [Eqs. (3) to (6)] have to be minimized. This can be done by using a normalization scheme, as used

in the previous studies ([12,13]). The normalized variables utilized in the numerical approach were as follows ([13,14]):

$$K^*=\frac{K_I}{K_{IC}};\quad J=\frac{A}{a_{ft}}t;\quad C^*=\frac{a}{a_{ft}};\quad \sigma^*=\frac{\sigma}{S_i};$$
$$\sigma^*_{\max}=\frac{\sigma_{\max}}{S_i};\quad \dot{\sigma}^*=\frac{\sigma^*}{J} \tag{7}$$

where K^*, J, C^*, σ^*, $\sigma^*_{\max}$, and $\dot{\sigma}^*$ are, respectively, normalized stress intensity factor, normalized time, normalized crack size, normalized applied stress, normalized maximum applied stress (in cyclic loading), and normalized stress rate. a_{ft} is the critical crack size in the inert condition, or is the initial crack size. Using these variables, the crack propagation rate of Eq. (2) yields

$$\frac{dC^*}{dJ}=[K^*]^n \tag{8}$$

The normalized SIF K^*, in constant stress-rate and cyclic (sinusoidal) stress testing can also be expressed, respectively ([13,14])

$$K^*=\dot{\sigma}^* J[C^*]^{1/2} \tag{9}$$

$$K^*=\left\{\frac{1+R}{2}+\frac{1-R}{2}\sin\left[\left(\frac{\omega a_{ft}}{A}\right)J\right]\right\}\sigma^*_{\max}[C^*]^{1/2} \tag{10}$$

where R is the stress (or load) ratio, defined as $R=\sigma_{\min}/\sigma_{\max}$ with $\sigma_{\min}$ being the minimum applied stress in cyclic loading, and ω is the angular velocity. The normalized SIF for constant stress loading is simply reduced to the case with both $R=1.0$ and $\sigma^*_{\max}=\sigma^*$ in Eq. (10).

The differential equation, Eq. (8), together with Eqs. (9) and (10), was solved numerically using a fourth-order Runge-Kutta method for a given loading history. The initial condition was $C^*=1.0$ at $J=0$. The instability conditions were $K^*=1.0$ and $dK^*/dC^*>0$. The solution in cyclic loading was independent of frequency ([14]); hence an arbitrary value of $\omega a_{ft}/A=100$ was used in the analysis. At interruption time, J_{int}, the corresponding variables including crack size, stress, stress intensity factor and time were provided to the next (second) loading sequence so that the required variables were determined until an instability condition was reached. Included in the typical input data were n, R, $\dot{\sigma}^*_f$ and $\sigma^*_{\max}$. The major assumption in the analysis was that only one mechanism, i.e., slow crack growth, was associated with failure.

Results of Numerical Solutions

Strength. Normalized strength (σ^*_f) as a function of percent of interruption time ($\varphi=J_{\text{int}}/J_f$) for the Case I loading history, is shown in Fig. 2. Six different values of n ranging from $n=5$ to 160 were employed. For each n value, the initial slow stress rate of $\dot{\sigma}^*=1\times10^{-5}$ was followed by the second loading, which was chosen as $\dot{\sigma}^*=1\times10^{-4}$, 1×10^{-3}, 1×10^{-2} and 1×10^{-1}. The choice of this range of $\dot{\sigma}^*$ was based on the typical range of applied stress rates commonly used in the actual constant stress-rate testing ([1]). For the case of $n=5$ where SCG susceptibility is high, the strength degradation with respect to the strength at $\varphi=0$ depends on interruption time, particularly with increasing stress rate. The maximum strength degradation of about 16 percent and 10 percent occurred at $\varphi=90$ percent, respectively, for the highest stress rate of $\dot{\sigma}^*=1\times10^{-1}$ and for the lowest rate of $\dot{\sigma}^*=1\times10^{-4}$. This indicates that the initial loading up to $\varphi=90$ percent resulted in somewhat appreciable crack growth/damage-accumulation. For $n=10$, the maximum strength degradation was about 3 percent both at $\dot{\sigma}^*=1\times10^{-1}$ and 1×10^{-4}. For the case of higher resistance to SCG, $n\geqslant20$, the strength degradation was negligible with less than 0.6 percent at $\varphi=90$ percent, indicating that crack growth/damage-accumulation rarely occurred during the first loading sequence. Therefore, it is concluded that for $n\geqslant20$ which is the case for most silicon nitrides and silicon carbides at elevated temperatures, the first loading sequence would not have any significant influence on crack growth/damage, leading to negligible strength degradation. The second loading rate, which is at least one order of magnitude greater than the first one, controls exclusively the strength via crack growth. An analysis on how a crack grows under a given loading history will be presented in a later section.

The results of strength as a function of percent of interruption time ($=\varphi$) for the Case II loading history, a combination of constant stress and constant stress-rate testing, Fig. 1(b), is presented in Fig. 3. Two to three different normalized applied stresses, ranging from $\sigma^*=0.2$ to 0.95 depending on n value, were used for each n value. Since the susceptibility to SCG decreases with increasing n, higher applied stress with narrow range was employed for higher n value. Two test rates of $\dot{\sigma}^*=1\times10^{-3}$ and 1×10^{-1} were used in the second loading sequence. As in the Case I loading, for $n\leqslant10$, strength degradation was significant to interruption time, particularly at both lower applied stress and higher stress rate of $\dot{\sigma}^*=1\times10^{-1}$. For $n\geqslant20$, strength degradation for a given value of φ was independent of either applied constant stress (σ^*) or applied stress rate. However, the overall degree of strength degradation occurring for the whole range of n's was greater in the Case II loading than in Case I loading. For $n=5$, the maximum strength degradation of 42 percent occurred at $\varphi=90$ percent with a loading combination of $\sigma^*=0.2$ and $\dot{\sigma}^*=1\times10^{-1}$. For $n=10$, about 20 percent strength degradation was observed at $\varphi=90$ percent for the combination of $\sigma^*=0.3$ and $\dot{\sigma}^*=1\times10^{-1}$. For $n\geqslant20$, the maximum strength degradation of 10 percent, 5 percent, 3 percent and 1 percent took place, independent of $\dot{\sigma}^*$, for $n=20$, 40, 80 and 160, respectively. More dependency of strength degradation on interruption time (φ), compared with the Case I loading, implies that the first, constant-stress loading sequence resulted in more crack growth damage-accumulation, thus leading to lower strength when the damaged specimen was subjected to the second constant stress-rate loading sequence.

The results of strength as a function of interruption time (φ) for the Case III loading, a combination of cyclic stress and constant stress-rate loading, see Fig. 1(c), is shown in Fig. 4. The second constant stress-rate loading sequence was preceded by the first, sinusoidal cyclic stress with a R ratio of $R=0.1$, until the specimen failed. Two to three different normalized maximum applied stresses, ranging from $\sigma^*_{\max}=0.2$ to 0.95, were used for each n value. Note that $\sigma^*_{\max}$ applied in the Case III loading was identical in magnitude to σ^* applied in the Case II loading. As in the case II loading, two stress rates of $\dot{\sigma}^*=1\times10^{-3}$ and 1×10^{-1} were also used in the second loading sequence. Comparing the results in Figs. 3 and 4, it can be readily evident that for the given n and $\sigma^*_{\max}=\sigma^*$, strength degradation in the Case III loading was almost the same as that in the Case II loading. It should be noted that constant stress (i.e., $R=1.0$) results in much shorter life than cyclic stress with $R=0.1$ ([11,14]). However, in terms of strength degradation as a function of *percent* of interruption time, either constant stress (Case II loading) or cyclic stress (Case III loading) yielded the similar result.

Slow Crack Growth. Typical examples of crack growth/damage-accumulation subjected to the three different loading histories are presented in Fig. 5 for both $n=10$ and 20. The figures show how an initial crack grows with time during the whole loading history until failure occurs. The loading combinations employed for each n were as follows: $\dot{\sigma}^*=1\times10^{-5}$ (first)/$\dot{\sigma}^*=1\times10^{-1}$ for Case I loading; $\sigma^*=0.5$ (first)/$\dot{\sigma}^*=1\times10^{-1}$ (second) for Case II loading; $\sigma^*_{\max}=0.5$ (first)/$\dot{\sigma}^*=1\times10^{-1}$ (second) for Case III loading. Note again that the case of $\varphi=1.0$ represents only the first loading sequence applied.

(a) *For* $n=10$ [Fig. 5A]. For the Case I loading [Fig. 5A(a)], an initial crack subjected to only the first loading sequence of

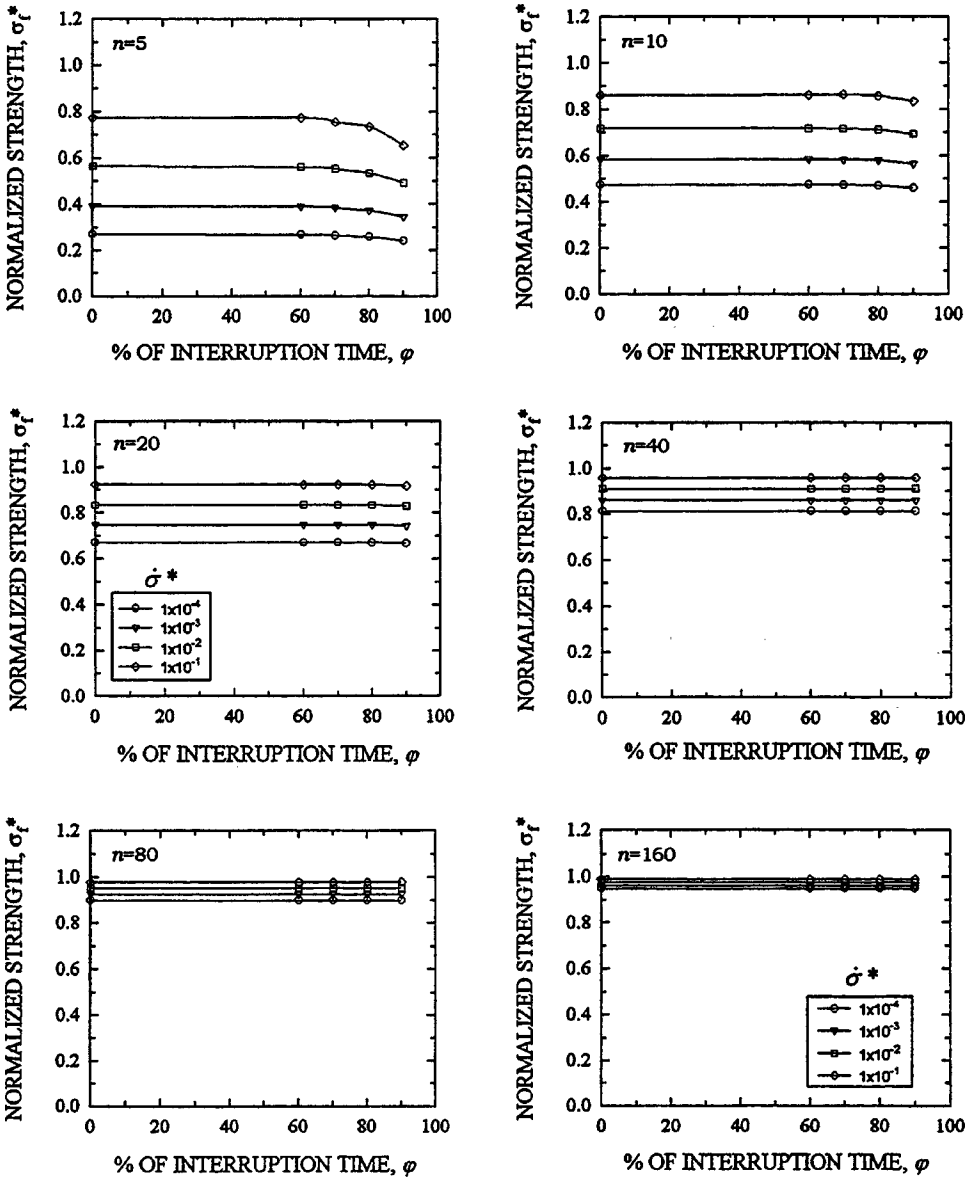


Fig. 2 Numerical results of normalized strength (σ_f^*) as a function of percent of interruption time (φ) for different values of slow crack growth (SCG) parameter n in Case I loading

$\dot{\sigma}^* = 1 \times 10^{-5}$ (i.e., $\varphi = 1.0$) remained almost unchanged in size for a long time, but started to grow very quickly at $J \geq 0.3848 \times 10^5$ until failure time of $J_f = 0.3849 \times 10^5$. This indicates that the initial crack started to grow to instability at a time greater than 95 percent of failure time. Therefore, any interruption of loading below $\varphi = 95$ percent did not give any significant crack growth so that the resulting strength after the second loading sequence remained unchanged (compared with the strength at $\varphi = 0$), irrespective of interruption time. This is also reflected as an insignificant strength degradation with respect to the strength at $\varphi = 0$, as shown in Fig. 2 for $n = 10$. Similar behavior as in the Case I loading was also observed in the Case II loading [see Figs. 5A(b)]. Most major crack growth occurred close to and/or at failure time. However, during the first static loading sequence, an initial crack started to grow earlier and greater in size than that of the Case I loading. Hence, the resulting critical crack size after the second loading sequence of $\dot{\sigma}^* = 1 \times 10^{-1}$ was increased with increasing interruption time (φ). As a result, strength degradation as a function of φ became much greater, compared with that of the Case I loading (see Fig. 3 for $n = 10$). For the Case III loading, crack growth behavior in terms of interruption time was almost identical to that of the Case II loading, as can be seen by comparing Fig. 5A(b) with Fig. 5A(c). The only difference between Case II and III loading lied in time to failure: Time to failure for a given n is always greater in cyclic ($R = 0.1$) than in constant stress ($R = 1.0$) loading ([11,14]). Therefore, the resulting strength as a function of *percent* of interruption time (φ) remained the same for either Case II (static) or Case III (cyclic) loading history, as shown in Figs. 2 and 3 for $n = 10$.

(b) *For* $n = 20$ (Fig. 5B). For the case of $n = 20$, the overall trend in crack growth behavior was very similar to the case of $n = 10$. Note again that an initial crack started to grow close to and/or at failure time, independent of the type of loading history. However, because of higher resistance to SCG in case of $n = 20$ the critical crack size at instability between $\varphi = 0$ and 0.9 was all smaller (and less dependant on interruption time) than that of the case for higher SCG susceptibility with $n = 10$. As a consequence, strength degradation as a function of interruption time (φ) was less significant compared with that of $n = 10$ (see Figs. 2 through 4 for $n = 20$).

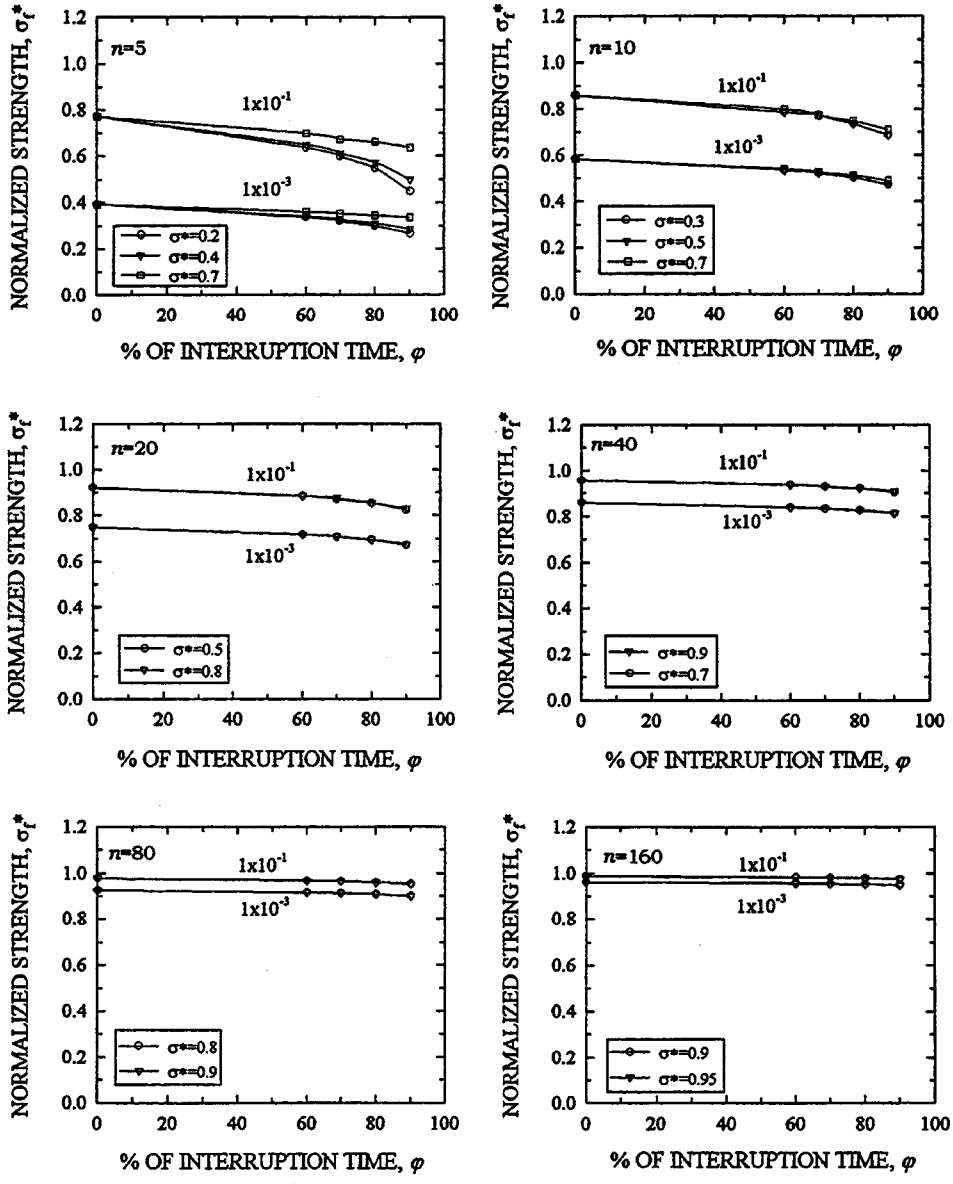


Fig. 3 Numerical results of normalized strength (σ_f^*) as a function of percent of interruption time (φ) for different values of slow crack growth (SCG) parameter n in Case II loading

Based on the numerical results on strength and crack growth, it can be summarized that strength degradation due to crack growth or damage accumulation as a result of the first loading sequence depends on interruption time φ and SCG parameter n for a given loading history. The strength degradation or degree of crack growth/damage-accumulation as a function of interruption time was significant for lower SCG parameters $n \leqslant 10$, but became insignificant with increasing SCG parameter of $n \geqslant 20$. This trend was observed more dominant for the Case I loading than the Case II or III loading history. The key factor that governs such strength degradation or crack growth behavior was that an initial crack started to grow typically close to and/or at failure time after a substantially long incubation time. This long incubation time, unique to ceramic materials exhibiting $n \geqslant 20$, was also a basis of the accelerating test methodology in constant stress-rate testing, where depending on n value appropriate preloading can be applied to test specimens prior to testing, thus saving a significant amount of test times ([15]).

Experimental Procedure

In order to verify the numerical solutions, experiments to cover different loading histories as specified in Fig. 1, were conducted at elevated temperatures. The nominal dimensions of rectangular-beam test specimens in accordance with test method ASTM C-1211 ([16]) were 3 mm by 4 mm by 50 mm, respectively, in height, width, and length. Test specimens were subjected to appropriate flexural loading depending on the type of loading history using SiC four-point flexure fixtures with 20-mm inner and 40-mm outer spans via electromechanical and servo-hydraulic test frames (Instron Models 8562 and 8501). All test specimens were equilibrated at test temperatures for about 20 min prior to testing. Four different materials including 96 wt percent alumina, NC132 silicon nitride, AS800 silicon nitride, and Hexoloy silicon carbide were used in the Case I loading, while only 96 wt percent alumina was used in the Case II loading. The reason for the choice of alumina in both Case I and II testing was that unlike other mate-

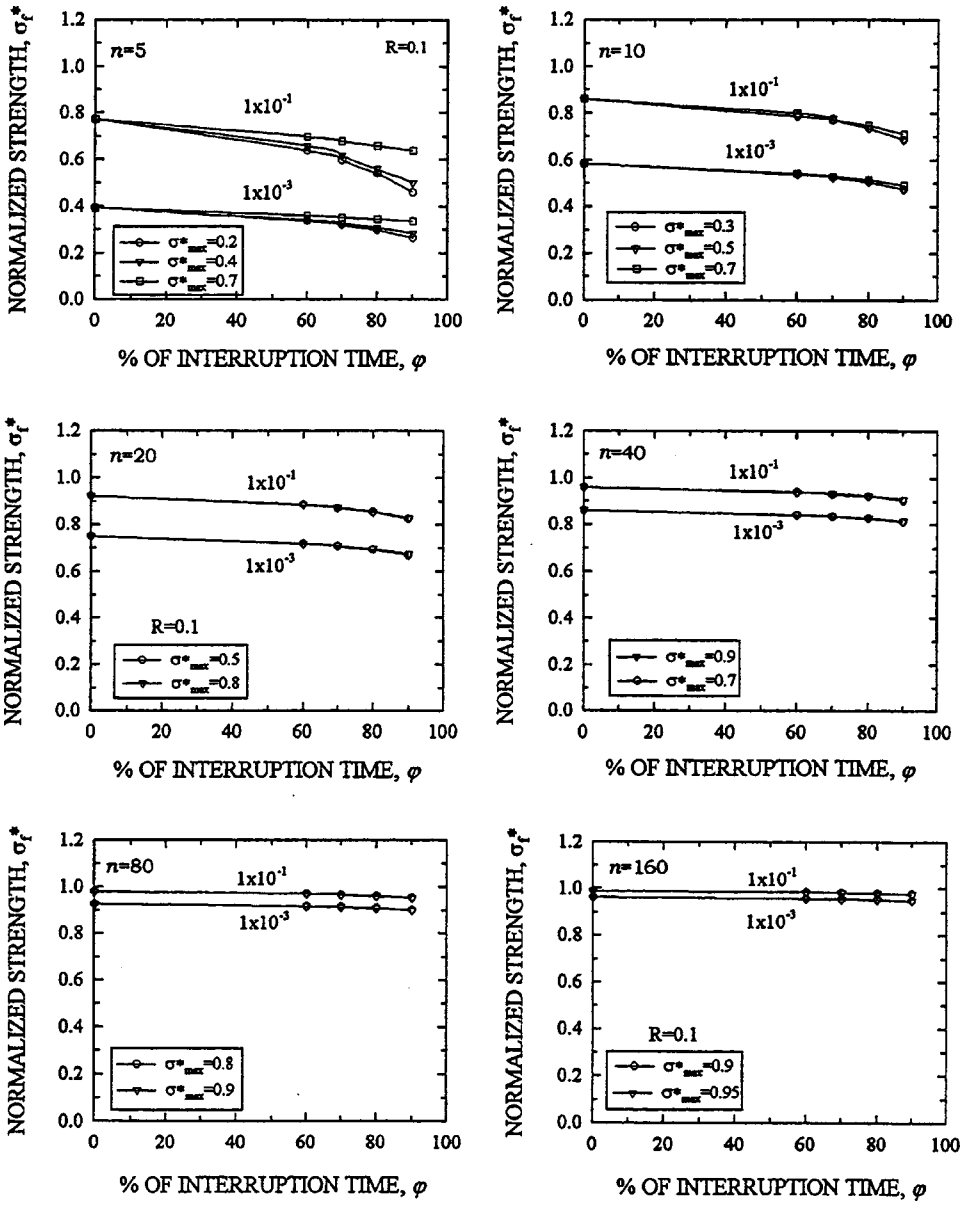


Fig. 4 Numerical results of normalized strength (σ_f^*) as a function of percent of interruption time (φ) for different values of slow crack growth (SCG) parameter n in Case III loading

rials, 96 wt percent alumina has exhibited a considerably small strength scatter with a Weibull modulus typically greater than 20 at either ambient or elevated temperatures ([6]). Hence, it would be possible to see material's response to life and strength more clearly and accurately with even a small number (about 5 at each condition) of test specimens. Also note that the alumina was very susceptible to SCG at elevated temperatures $\geqslant$800°C with significantly low values of SCG parameter of $n=7-12$ ([17]), so that it would be much easier using the alumina to scrutinize the influence of SCG/damage-accumulation on the combined loading sequences more accurately. The experimental work for the Case III loading was not conducted in this paper, primarily due to limited availability of test specimens.

(a) Case I Loading. In the Case I load testing, the loading history included a slow test rate of 0.033 MPa/s for the first loading sequence and then a fast test rate of 33.33 MPa/s for the second loading sequence. The percentage of interruption time ($\varphi=t_{\text{int}}/t_f$) ranged from $\varphi=70$ to 90 percent. The average failure time ($=t_f$) of test specimens only subjected to the first loading sequence (with 0.033 MPa/s) was determined from the previous studies ([6,18]), and used here as a reference value to calculate t_{int} for a given value of φ. Four ceramics including 96 wt percent alumina NC132 silicon nitride, AS800 silicon nitride, and Hexoloy silicon carbide were tested at temperatures of 1000, 1100, 1200 and 1371°C, respectively. Typically a total of five specimens, depending on material, were used at each value of φ. The major mechanical and physical properties of the test materials such as Young's modulus, density, fracture toughness, strength, and slow crack growth can be found elsewhere ([18]).

(b) Case II Loading. Constant stress (''static fatigue'') testing for 96 wt percent alumina was first conducted in flexure at 1000°C to determine SCG behavior and thus to obtain the time-to-failure data. Four different applied stresses ranging from 50 to 100 MPa were employed, with a total of five to nine specimens tested at each applied stress. The Case II loading history consisted of a constant stress (for the first loading sequence) and a fast stress rate of 33.33 MPa/s (for the second sequence). Two applied

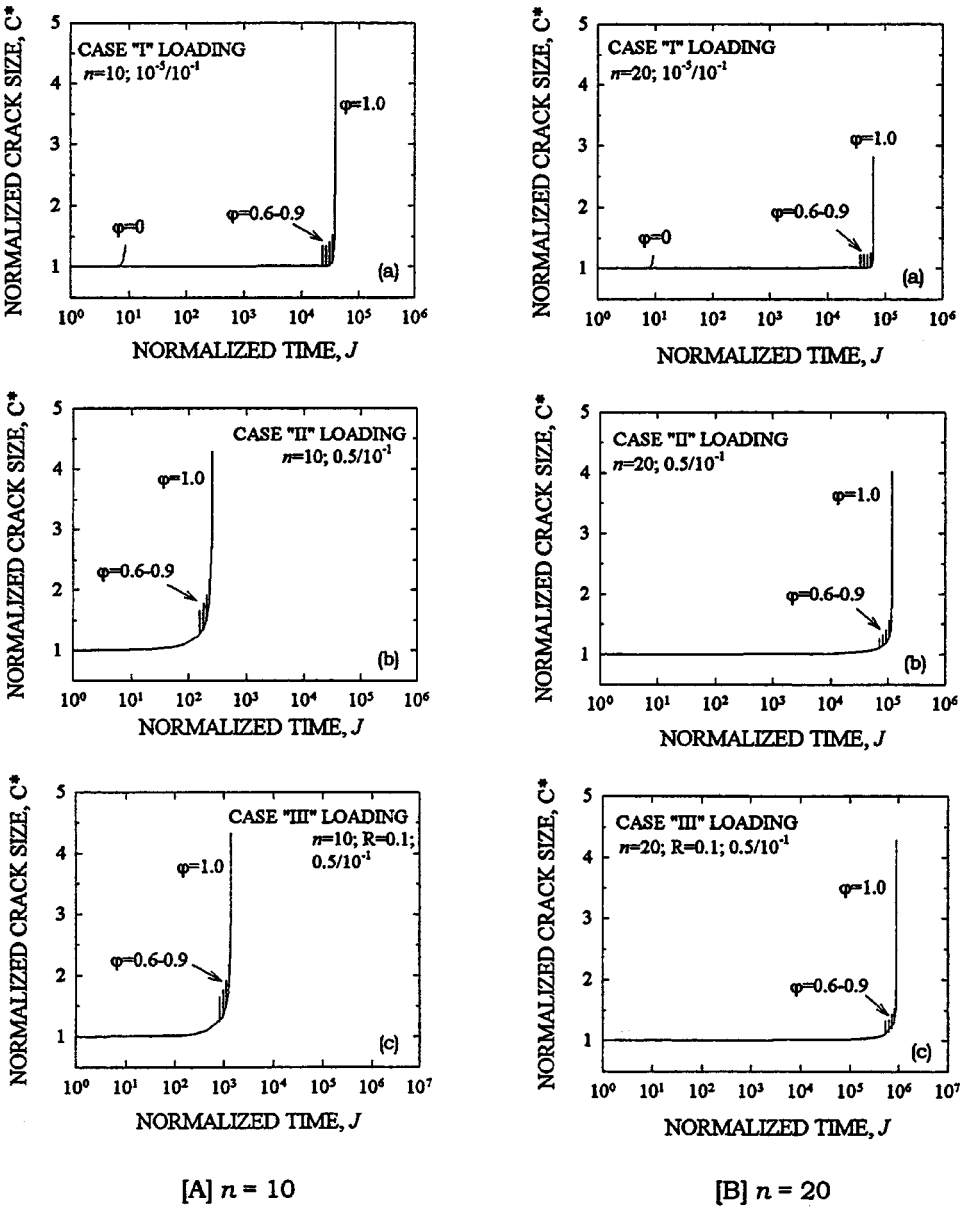


Fig. 5 Numerical results of normalized crack size (C^*) as a function of time (J) for different values of interruption time (φ) in three loading histories: [A] for SCG parameter $n=10$; [B] for SCG parameter $n=20$

stresses of 50 and 65 MPa were used in the first loading sequence. Three different values of interruption time, $\varphi=60$, 75, and 90 percent, were utilized at each applied stress, with a total three to five specimens tested at each interruption time.

Experimental Results and Discussion

(a) Case I Loading. The results for the Case I loading tests for 96 wt percent alumina, NC132 silicon nitride, AS800 silicon nitride, and Hexoloy silicon carbide are summarized in Fig. 6. The figure included flexure strength as a function of percent of interruption time φ for each material. The horizontal line represents the strength determined with zero interruption time $\varphi=0$, that is, the ''fast''-fracture strength evaluated at 33.33 MPa/s ([6,18]). The three materials including NC132 and AS800 silicon nitrides and Hexoloy silicon carbide exhibited a somewhat appreciable variation (in average sense) in strength between $\varphi=0$ and $\varphi=80$ or 90 percent. It is believed that this was attributed to the inherently large strength scatter, typical of advanced ceramics that ranges commonly from 10 to 15 in Weibull modulus. By contrast, 96 wt percent alumina exhibited a very small scatter.

(b) Case II Loading. Figure 7 shows the results of constant stress testing for 96 wt percent alumina at 1000°C. The slow crack growth parameters n and D_s in Eq. (4) were determined as $n=9.8$ and $D_s=4.69\times10^{20}$ with units of ''MPa'' in stress and ''second'' in time. Note that SCG parameter n determined from constant stress testing was in reasonable agreement with $n=8.3$ from constant stress-rate testing determined from a previous study ([6]). The results of the Case II loading tests was presented in Fig. 8, where strengths determined at 33.33 MPa/s, after the first loading sequence of constant stress of 50 or 65 MPa, was plotted as a function of percent of interruption time (φ). As seen in the figure, the strength exhibited a significant scatter particularly at $\varphi=75$ and 90 percent, much greater than that exhibited in the Case I

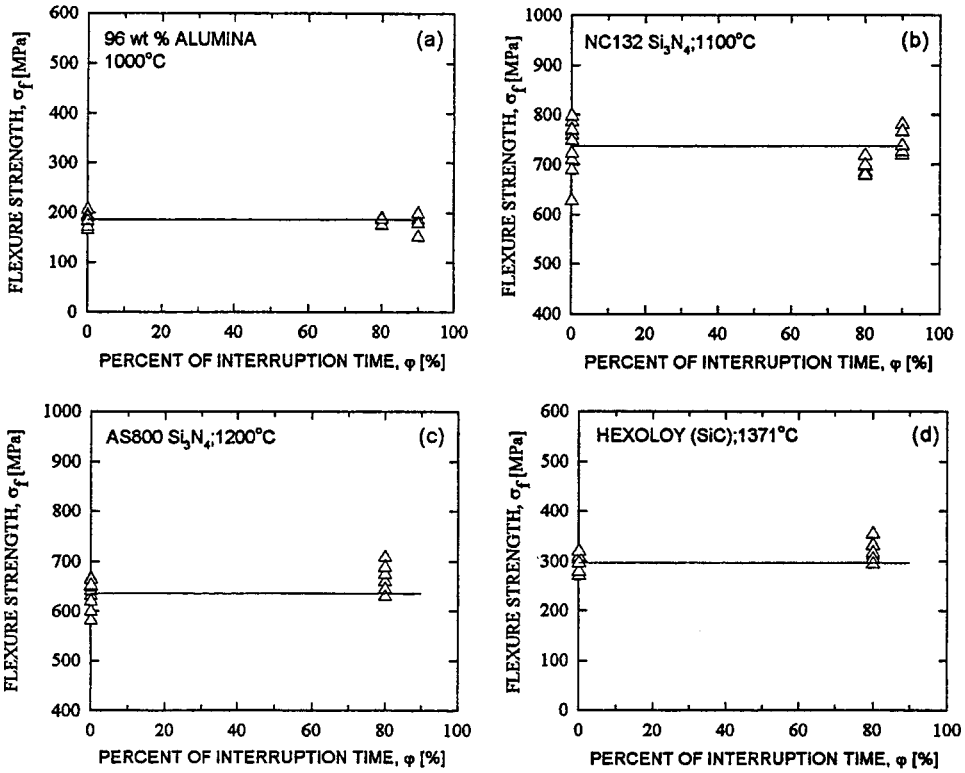


Fig. 6 Experimental results of flexural strength as a function of percent of interruption time (φ) for Case I loading tests, determined from 96 wt percent alumina, NC132, and AS800 silicon nitrides, and Hexoloy silicon carbide at elevated temperatures. Each solid line represents the mean strength at $\varphi=0$.

loading history for the same alumina material. It is believed that this was attributed to the fact that no exact failure time of each individual test specimen subjected to the Case II loading could be known and that as a result the actual corresponding interruption time for each test specimen could not be determined. This will be discussed in a later section.

(c) Comparison of Experimental Data With Numerical Solutions. The comparison of strength as a function of φ between the experimental data and the numerical solutions for each loading history was made and presented in Figs. 9 and 10. The *reduced* strength (σ_r^*) used here was defined such that strength determined at any given value of φ was normalized with respect to the strength determined at $\varphi=0$, which is expressed as follows:

$$\sigma_r^* = \frac{\sigma_\varphi}{\sigma_{\varphi=0}} \tag{11}$$

where σ_φ is the strength at any given value of φ (which is determined at a fast test rate of 33.33 MPa/s after the first loading sequence) and $\sigma_{\varphi=0}$ is the strength determined at $\varphi=0$ (which is simply the fast-fracture strength determined at 33.33 MPa/s without any first loading sequence).

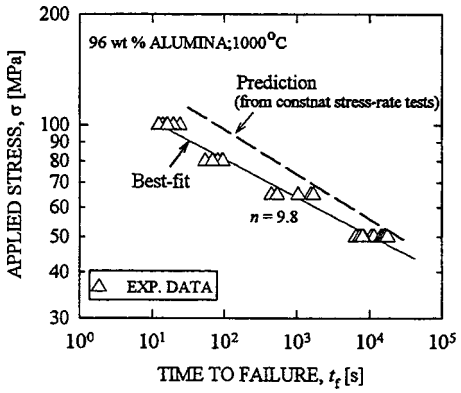


Fig. 7 Experimental results of constant stress (''static fatigue'') testing for 96 wt percent alumina at 1000°C. A prediction made from the constant stress-rate (''dynamic fatigue'') testing data ([6]) was included as a dotted line.

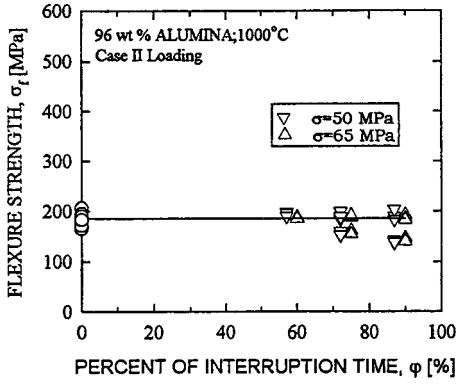


Fig. 8 Experimental results of flexure strength as a function of percent of interruption time (φ) for Case II loading tests, determined from 96 wt percent alumina at 1000°C. Two different constant stresses of 50 and 65 MPa were employed. The solid line represents the mean strength at $\varphi=0$.

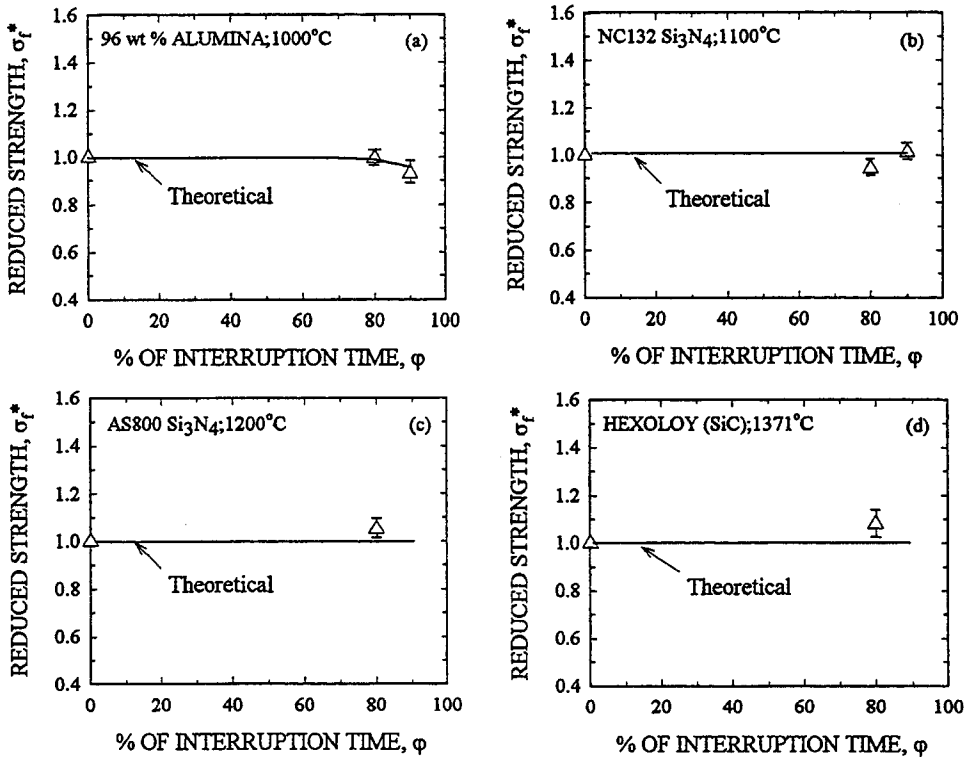


Fig. 9 Comparison of experimental data with numerical solutions (''theoretical'') in Case I loading: (a) 96 wt percent alumina, (b) NC132 silicon nitride, (c) AS800 silicon nitride, and (d) Hexoloy silicon carbide. Error bars represents ±1.0 standard deviation, normalized with respect to a mean strength at $\varphi=0$.

(i) Case I Loading. As seen in Fig. 9, except for 96 wt percent alumina, the discrepancy between the experimental mean-strength data and the numerical solutions was somewhat large. However, as aforementioned, considering low Weibull modulus (10–15) typical of many advanced ceramics including the test materials (except 96 wt percent alumina), the discrepancy is believed rather statistically insignificant. The experimental data for 96 wt percent alumina which exhibited a considerably small strength scatter were in good agreement with the theoretical prediction. Based on the results shown in Fig. 9, several conclusions were made. First, the governing mechanism associated with failure for these test materials was slow crack growth. Some materials such as alumina and NC132 silicon nitride exhibited some degree of creep deformation ($\leq$0.2 percent creep strain) at a lower test rate of 0.033 MPa/s. Despite such creep mechanism presented, the agreement between experimental data and prediction was still reasonable, implying SCG to be a dominant failure mechanism. Second, the phenomenon typical of advanced ceramics, which showed numerically that for $n \geq 10$ in constant stress-rate condition an initial crack started to grow at and/or close to failure time after a long incubation time, was verified. This was verified from the results that percent of interruption time up to $\varphi=90$ percent did not show any significant crack growth or damage accumulation, as reflected in insignificant strength degradation. This ''long-incubation'' phenomenon was also validated previously by the accelerating test technique developed in constant stress-rate testing ([15]). Finally, the slow crack growth formulation of Eq. (2) well described the actual SCG behavior of the test materials at elevated temperatures. Note that the numerical solution was made exclusively based on such SCG formulation so that if poor agreement would exist, then it would be indicative of inapplicability of the SCG formulation to the actual material behavior.

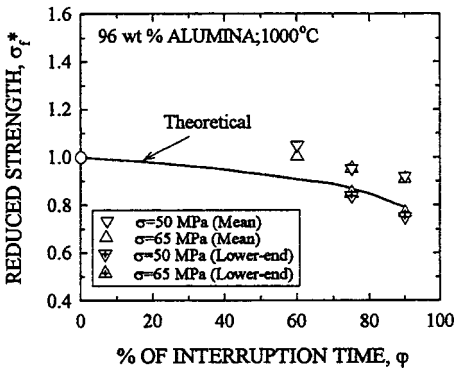


Fig. 10 Comparison of experimental data with numerical solutions (''theoretical'') in Case II loading tests for 96 wt percent alumina at 1000°C. Each data point with ''triangle'' symbol represents an overall mean strength value (normalized). The data points with triangle's symbols (inner cross-marked) represent the lower-end strength values (normalized) from the data in Fig. 8. Error bars were omitted from plots for clarity.

(ii) Case II Loading. The *reduced* strength as a function of percent of interruption time (φ) for 96 wt percent alumina is presented in Fig. 10. Unlike the Case I loading history (Fig. 9), the difference between the experimental mean-strength data (with ''triangle'' symbols) and the theoretical prediction was amplified even for the same alumina material. The reason for this discrepancy can be reasoned as follows. Because of the two combined loading sequences, no exact failure time of each individual test specimen exclusively subjected to the first loading sequence (constant stress) could be known so that the actual interruption time

for each test specimen can never be determined. Note that the (nominal) time-to-failure at each applied stress was taken as an average failure time determined from the specimens subjected to 50 MPa or 65 MPa. The corresponding interruption time was calculated simply from a relation of $t_{\text{int}}=\varphi t_f$ for a given value of φ. Furthermore, as seen in Fig. 7, the scatter in time to failure was greater than that of strength typically observed in constant stress-rate testing ([6]), thus further increasing the uncertainty in failure time. Therefore, some specimens would have been actually subjected to greater interruption time than the nominal interruption time, while other specimens to less interruption time. As a consequence, this would have resulted in difference in SCG damage accumulation even with the same nominal φ, thereby resulting in a wide scatter in strength. The discrepancy would be small if the strength at lower end of data points, which would be close to the actual failure-time data, was used, as shown in the figure with the ''circle'' symbol. Certainly, the discrepancy will be diminished if a large number of test specimens are used.

(d) Comparison of SCG Behavior Between Constant Stress-Rate and Constant Stress Testing. As stated before, the SCG parameter n for 96 wt percent alumina was in reasonable agreement between constant stress-rate and constant stress testing with the respective values of $n=8.3$ and 9.8. It is possible to convert the SCG data from one test method to another by using the appropriate relations shown in Eqs. (3) to (6). The resulting comparison is depicted in Fig. 7, where the prediction from constant stress-rate (''dynamic fatigue'') ([6]) to constant stress (''static fatigue'') testing was included as a dotted line. Considering the inherent scatter in time to failure, reasonable agreement was found between the two data, implying that the mechanism associated with failure for both cases was presumably slow crack growth. It was observed that creep deformation in constant stress testing was much greater than that in constant stress-rate testing, since test time was much longer in constant stress testing. A maximum creep strain of about 0.4 percent was found for the test specimens subjected to an applied stress of 50 MPa. In spite of such appreciable creep deformation, overall agreement between the two test methods was reasonable, again indicative of SCG as an operative failure mechanism involved in both constant stress-rate and constant stress testing for this material system. However, some effect by creep deformation, particularly in constant stress (''static fatigue''), should not be overlooked, since some stress redistribution would be expected for test specimens subjected to long-term, constant stress testing. This creep effect, of course, becomes more dominant at higher temperatures, resulting in more deviation between the two test methods.

Since the predictions of strength were in reasonable agreement with experimental data, independent of the type of loading history or the type of testing, it is concluded that slow crack growth was a unique mechanism associated with failure for 96 wt percent alumina as well as for other test materials. Therefore, it is feasible in principle using the numerical analysis developed in this paper to predict life and/or strength degradation for any given simple or complex loading history as long as an explicit mathematical expression of load history can be made. More importantly, the analysis can be used in conjunction with appropriate experiments—using one of three loading histories—to promptly assess a governing failure mechanism involved in the conventional life-prediction testing, which could be either constant stress-rate, constant stress, or cyclic stress testing. Furthermore, the analysis would be possible to be used at least as a quantitative tool for damage assessment by estimating crack growth/damage-accumulation of a structural component in service. A verification run for the numerical analysis for the Case III loading history was not made in this paper so it will be a subject of future study.

Conclusions

(1) For $n\leqslant 10$, simulated strength degradation as a function of percent of interruption time was significant for all the Cases I, II and III loading histories. For the given interruption time φ and n, the degradation was greater in constant or cyclic stress (Case II or III) than in constant stress-rate loading (Case I). By contrast, for $n\geqslant 20$, the degradation became negligible either for the Case I, II, or III loading history.

(2) The numerical solutions of strength degradation were examined using the experimental data determined at elevated temperatures from four different advanced ceramics—two silicon nitrides, one silicon carbide and one alumina—for the Case I loading history, and from alumina for the Case II loading. The experimental data was in reasonable agreement with the numerical solutions for both loading histories. Also the phenomenon typical of advanced ceramics, which has shown that in constant stress-rate condition an initial crack starts to grow at and/or close to failure time after a long incubation time, was verified.

(3) The numerical analysis that assumed *slow crack growth* as a governing failure mechanism was in good agreement with the experimental data. Despite some degree of creep deformation presented, slow crack growth was presumably a significantly operating failure mechanism involved in all the test materials, regardless of type of loading history or type of test method. Notwithstanding some deviation, there was also reasonable agreement in SCG behavior between constant stress-rate testing and constant stress testing for 96 wt percent alumina. This supports that the widely utilized SCG formulation of $v=A[K_I/K_{IC}]^n$ described reasonably the slow crack growth behavior of the test materials.

(4) It appears that the analysis in conjunction with proper experiments, using one of three loading histories (Case I preferred), may be utilized as a tool for damage (SCG and creep) assessment of test specimens subjected to conventional life prediction testing. However, care must be exercised when several failure mechanisms such as SCG, creep and oxidation are actively operating in series at higher temperatures. A convenient analytical tool(s) accessible to such combined conditions, currently not available though, is inevitable.

Acknowledgments

The authors are thankful to R. Pawlik of NASA Glenn for the experimental work. This work was sponsored in part by the HITEMP project, NASA Glenn Research Center, Cleveland, OH.

References

[1] ASTM C1368-97, 1998, ''Standard Test Method for Determination of Slow Crack Growth Parameters of Advanced Ceramics Using Constant Stress-Rate Flexural Testing at Ambient Temperature,'' *Annual Book of ASTM Standards*, Vol. 15.01, ASTM, Philadelphia, PA.

[2] ASTM C1465-00, 2000, ''Standard Test Method for Determination of Slow Crack Growth Parameters of Advanced Ceramics Using Constant Flexural Stress-Rate Testing at Elevated Temperatures,'' *Annual Book of ASTM Standards*, Vol. 15.01, ASTM, Philadelphia, PA.

[3] Quinn, G. D., and Morrell, R., 1991, ''Design Data for Engineering Ceramics: A Review of the Flexure Test,'' J. Am. Ceram. Soc., **74**, pp. 2037–2066.

[4] Fett, T., and Munz, D., 1985, ''Determination of Crack Growth Parameter N in Ceramics Under Creep Condition,'' J. Test. Eval., **13**, pp. 143–151.

[5] Jadaan, O. M., 1991, ''Life Prediction for Ceramic Tubular, Components,'' in *Life Prediction Methodologies and Data for Ceramic Materials*, ASTM STP 1201, edited by Brinkman, C. R., and Duffy, S. D., American Society for Testing and Materials, Philadelphia, pp. 309–332.

[6] Choi, S. R., and Gyekenyesi, J. P., 1998, ''Some Limitations in the Elevated-Temperature, Constant Stress-Rate Flexural Testing for Advanced Ceramics with Reference to the New, Ambient-Temperature Test Standard ASTM C1368,'' Ceram. Eng. Sci. Proc., **19**, pp. 595–605.

[7] Chuck, L., McCullum, D. E., Hecht, N. L., and Goodrich, S. M., 1991, ''High Temperature Tension-Tension Cyclic Fatigue for a Hipped Silicon, Nitride,'' Ceram. Eng. Sci. Proc., **12**, pp. 1509–1523.

[8] Wiederhom, S. M., 1974, ''Subcritical Crack Growth in Ceramics,'' in *Fracture Mechanics of Ceramics*, Vol. 2, edited by Bradt, R. C., Hasselmann, D. P. H., and Lange, F. F., Plenum, New York, pp. 613–646.

[9] Ritter, J. E., 1978, ''Engineering Design and Fatigue Failure of Brittle Materials,'' in *Fracture Mechanics of Ceramics*, Vol. 4, edited by R. C. Bradt, D. P. H. Hasselmann, and F. F. Lange, Plenum New York, pp. 661–686.

[10] Evans, A. G., 1974, ''Slow Crack Growth in Brittle Materials under Dynamic Loading Conditions,'' Int. J. Fract., **10**, pp. 251–259.

[11] Evans, A. G., and Fuller, E. R., 1974, ''Crack Propagation in Ceramic Materials under Cyclic Loading Conditions,'' Metall. Trans., **5**, pp. 27–33.
[12] Lawn, B. R., Marshall, D. B., Anstis, G. R., and Dabbs, T. P., 1981, ''Fatigue Analysis of Brittle Materials Using Indentation Flaws, Part 1. General Theory,'' J. Mater. Sci., **16**, pp. 2846–2854.
[13] Choi, S. R., Ritter, J. E., and Jakus, K., 1990, ''Failure of Glass with Subthreshold Flaws,'' J. Am. Ceram. Soc., **72**, pp. 268–274.
[14] Choi, S. R., and Salem, J. A., 1996, ''Cyclic Fatigue of Brittle Materials with an Indentation-Induced Flaw System,'' Mat. Sci. Eng., **A208**, pp. 126–130.
[15] (a) Choi, S. R., and Gyekenyesi, J. P., 1997, ''Fatigue Strength as a Function of Preloading in Dynamic Fatigue Testing of Glass and Ceramics,'' ASME J. Eng. Gas Turbines Power, **119**, pp. 493–499. (b) Choi, S. R., and Salem, J. A., 1996, ''Preloading Technique in Dynamic Fatigue Testing of Glass and Ceramics with an Indentation Flaw System,'' J. Am. Ceram. Soc., **79**, pp. 1228–1232.
[16] ASTM C1211-98a, 1998, ''Standard Test Method for Flexural Strength of Advanced Ceramics at Elevated Temperatures,'' *Annual Book of ASTM Standards*, Vol. 15.01, ASTM, Philadelphia, PA.
[17] Choi, S. R., and Salem, J. A., 1998, ''Ultra'-Fast Fracture Strength of Advanced Ceramics at Elevated Temperatures,'' Mat. Sci. Eng., **A242**, pp. 129–136.
[18] Choi, S. R., and Gyekenyesi, J. P., 2000, ''Ultra'-Fast Fracture Strength of Advanced Structural Ceramics at Elevated Temperatures: An Approach to High-Temperature 'Inert' Strength,'' presented at the 7th Fracture Mechanics of Ceramics Symposium, July 20–23, 1999, Moscow, Russia: to be published in *Fracture Mechanics of Ceramics*, Plenum, New York.

Design, Fabrication, and Testing of Ceramic Joints for High Temperature SiC/SiC Composites

M. Singh
Dynacs Engineering Co.
NASA Glenn Research Center
Cleveland, OH 44135

E. Lara-Curzio
Metals and Ceramics Division
Oak Ridge National Laboratory
Oak Ridge, TN 37831

Various issues associated with the design and mechanical evaluation of joints of ceramic matrix composites are discussed. The specific case of an affordable, robust ceramic joining technology (ARCJoinT) to join silicon carbide (CG-Nicalon™) fiber-reinforced-chemically vapor infiltrated (CVI) silicon carbide matrix composites is addressed. Experimental results are presented for the time and temperature dependence of the shear strength of these joints in air up to 1200°C. From compression testing of double-notched joint specimens with a notch separation of 4 mm, it was found that the apparent shear strength of the joints decreased from 92 MPa at room temperature to 71 MPa at 1200°C. From shear stress-rupture testing in air at 1200°C it was found that the shear strength of the joints decreased rapidly with time from an initial shear strength of 71 MPa to a value of 17.5 MPa after 14.3 h. The implications of these results in relation to the expected long-term service life of these joints in applications at elevated temperatures are discussed. [DOI: 10.1115/1.1362664]

Introduction

Ceramic matrix composites (CMC's) are leading candidate materials for a wide variety of high temperature, high performance aerospace and land-based applications. The potential applications of CMC's in the aerospace industry include combustor liners, exhaust nozzles, and a number of other aircraft gas turbine and space propulsion components. The land-based applications of these materials include radiant burners, hot gas filters, high-pressure heat exchanger tubes, and combustor liners in industrial gas turbine engines. In addition, there are a number of potential uses of CMC's for the first wall and blanket components of nuclear fusion reactors.

At present, most approaches for the fabrication of ceramic matrix composites (CMC's) are based on the gas or liquid infiltration of the ceramic matrix into a ceramic fiber preform. Among the various technologies available for the manufacturing of CMC's, chemical vapor infiltration (CVI) is the most mature and most widely used. However, the main drawbacks of this process are the (1) difficulty in achieving uniform and complete densification of large and/or geometrically complex shaped components, and (2) lengthy processing time. Therefore, design considerations for the manufacture of large CVI ceramic matrix composite components and structures are being based on using technologies for joining/attaching smaller-sized components with simpler geometry. Thus, ceramic joining is today recognized as an enabling technology for the successful implementation of CMC's in some of the aforementioned applications ([1–4]).

Various joint design issues are discussed in this paper and the particular case of joints of silicon carbide (CG-Nicalon™) fiber-reinforced CVI-SiC composites is addressed. These joints were produced using the affordable, robust ceramic joining technology (ARCJoinT) ([5–12]). This technique is based on the reaction forming approach, and will be described and discussed in terms of global design considerations. This joining technology is unique in terms of producing joints with tailorable microstructures. The formation of joints by this approach is attractive since the thermomechanical properties of the joint interlayer can be tailored to be very close to those of the base materials. In addition, high temperature fixturing is not needed to hold the parts at the infiltration temperature. A variety of silicon carbide-based ceramics and fiber-reinforced composites have been joined using this approach ([5–12]). Figure 1 is a photograph showing components fabricated by joining fiber-reinforced silicon carbide matrix composite subelements using ARCJoinT.

Joint Design Issues

Among the various alternatives available to overcome the limitations of the many fabrication technologies for the manufacture of large CMC components and structures of complex shape, the joining of smaller components with simple geometry appears to be the most promising and practical. Although the use of joints for the manufacture of structural components will simplify the manufacturing process, it introduces additional complexities for the design and analysis of these structures. In addition to the approaches that are under consideration for the design, analysis and manufacture of single CMC parts, it will be essential to develop standardized tests, design methodologies, and life-prediction analyses for structures incorporating joints of these materials. Although substantial work has been conducted in the design and characteriza-

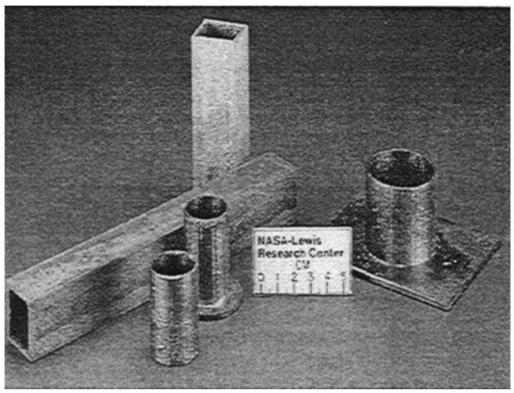


Fig. 1 Photograph showing components fabricated by joining fiber-reinforced silicon carbide matrix composite subelements

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Munich, Germany, May 8–11, 2000; Paper 00-GT-069. Manuscript received by IGTI Oct. 1999; final revision received by ASME Headquarters Oct. 2000. Associate Editor: D. Wisler.

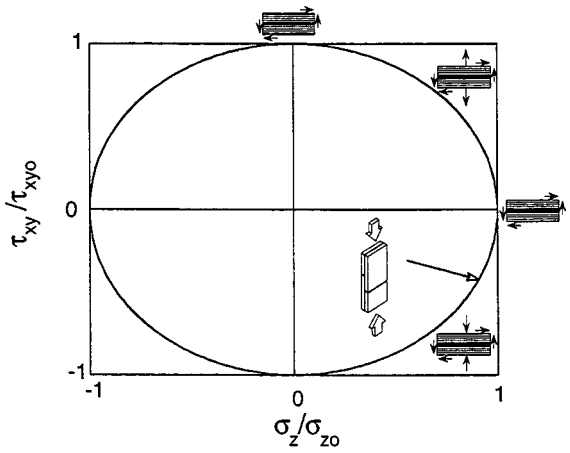


Fig. 2 Schematic of a joint design map where the shear and normal stresses have been normalized by the virgin isothermal joint strengths. The envelope defines the strength of the joint.

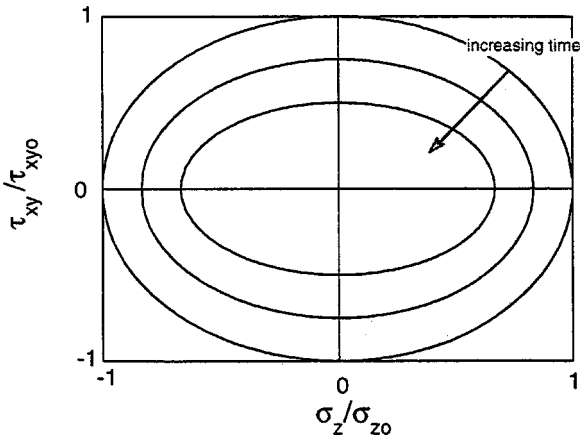


Fig. 4 Schematic of a joint design map where the design envelope shrinks due to a degradation of strength with time

tion of joints of fiber-reinforced polymer matrix composites and metals, comparable work for CMC's is limited because these materials are relatively new. For the ceramic-metal systems, various joint designs and design criteria have been established ([1–3]). These designs accommodate a number of factors including stresses and stress distribution in the joint regions, which are dependent upon joint configuration and chemical and thermal property mismatch between the joint and substrate materials. A wide variety of test methods ([1–3]) have also been used to determine the tensile strength, peel strength, flexural strength, shear strength, and compressive strength of ceramic-metal joints. Although, many of these tests could be implemented for the evaluation of joints of CMC's, work will still be needed to standardize these tests in order to address issues that are unique to ceramics and ceramic matrix composites.

Design considerations for ceramics and ceramic matrix composites are typically more complicated than for other conventional engineering materials, in view of the fact that in the majority of the cases these materials will be used at elevated temperatures, under stress, in aggressive environments, and for very long periods of time. Ideally, it would be desirable to evaluate the strength and reliability of joints under states of stress and/or strain that are comparable to those that will be found in the intended application and more often than not, service conditions involve multiaxial states of stress.

One approach for the design and analysis of reliability and durability of joints of CMCs that takes into account the multiaxial state of stress and/or strain that these joint specimens will face consists of using failure maps. Figure 2 shows a schematic of a failure map for loading conditions associated with a biaxial state of stress (i.e., combinations of shear and normal stresses) that are often found in the application of joints. This state of stress would be encountered for example, when joining tubes to a manifold in a heat exchanger (Fig. 3). For the purpose of designing joints of CMC's, it would be desirable to develop failure envelopes as a function of temperature and environment. Since most of the envisioned applications of CMC's, specially those for the energy industries, have service lives that are measured in several thousands of hours, it will also be necessary to generate information about how failure envelopes evolve as a function of time. Considering the aggressive service conditions that are associated with most of the envisioned applications of CMC's, it is likely that the strength of the joints will be controlled by degradation mechanisms that include: slow crack growth, stress rupture, creep, and oxidation and/or corrosion. Therefore, in their simplest form, (i.e., for applications where stresses would remain constant with time) failure maps will be defined by failure envelopes that shrink with time as a result of strength degradation. Figure 4 illustrates this idea.

In this paper, experimental results are presented for the virgin shear strength and the shear stress rupture behavior of joined silicon carbide (CG-Nicalon™) fiber-reinforced CVI SiC composites. The shear strength of the joints was determined by the compression of double-notched specimens. Although, there is a consensus on the lack of a good test for determining the shear strength of joined composites, the compression of double-notched specimens is a popular test because it requires relatively small samples[1] and it is relatively simple to conduct. In this test, shear failure is induced between two notches that are anti-symmetrically located on opposite sides of the specimen at an equal distance from the mid plane. The drawback of this test method is that because interlaminar shear failure is induced by the stress concentration at the root of the notches, the shear stress distribution between the notches is not uniform. Furthermore, in addition to shear stresses in the gauge section of the specimen, normal compressive stresses are also induced in this region. Therefore, in reference to the diagram in Fig. 2, data generated by the compression of double-notched specimens would lay in the third quadrant of the failure map in Fig. 2. This test method was recently standardized by ASTM for the determination of the interlaminar shear strength of one-dimensions (1-D) and 2-D continuous fiber-reinforced ceramic matrix composites at room temperature and elevated temperatures (ASTM C1292, ASTM C1425) and seems appropriate for the determination of the shear strength of joints ([13–15]). Figure 5 shows the schematic of the double-notched joint specimen.

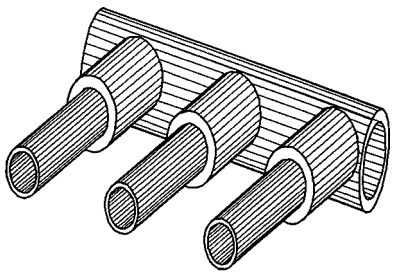


Fig. 3 Schematic of tubes joined to a manifold where stresses in the joint regions involve multiaxial states of stress. (e.g., normal plus shear stresses).

[1]Because monolithic ceramics and ceramic matrix composites exhibit volumetric dependence of strength, it will be necessary to appropriately scale the results from the evaluation of joints using ''small'' specimens, according to the specific application.

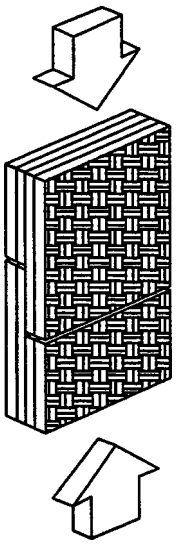


Fig. 5 Schematic of the double-notched shear specimen

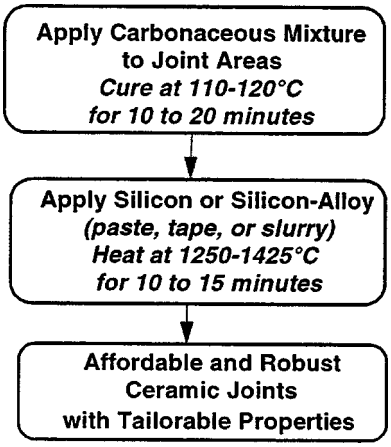


Fig. 6 Schematic of the process for joining of fiber-reinforced silicon carbide matrix composites

Fabrication

The composites used in this study were obtained from Honeywell Advanced Composites, Inc., Newark, DE. These composites were fabricated with plain weave fabric of SiC (CG Nicalon™) fibers[2] with a thin pyrolytic carbon fiber coating. The SiC matrix was deposited via chemical vapor infiltration, and the matrix contained an oxidation-resistance ''enhancement'' of proprietary composition. The fiber volume fraction was 44 percent. As-fabricated specimen surfaces were cleaned in acetone and dried before joining. A flow diagram of the affordable, robust ceramic joining technology (ARCJoinT) is given in Fig. 6. The joining process ([7,11,12]) begins with the application of a carbonaceous mixture in the joint area and a curing step at 110–120°C for 10–20 min while holding the items to be joined in a fixture thus fastening the pieces together. Then, a silicon-titanium alloy in paste form was applied around the joint region and heated to 1360°C for 15 min. The molten silicon-titanium alloy reacts with carbon to form silicon carbide with controllable amounts of silicon and titanium disilicide phases. By controlling the consistency of the carbonaceous paste and the magnitude of the applied force with the fixture, the thickness of the joint can thus be controlled. A wide variety of silicon-carbide-based ceramics and ceramic matrix composites, consisting of different sizes and shapes, have been joined using this technology (Fig. 1).

After the specimens were joined, double-notched specimens were machined to conform to the specimen geometry in Fig. 5 (6 mm wide, 22 mm long, 6.15 mm thick) using a diamond-impregnated wheel. The notches were machined in several passes with a removal rate of 2.5 μm/pass using a 0.5-mm-thick wheel. The final notch separation was 4.0 ± 0.05 mm. The compression tests of double-notched specimens were conducted in ambient air at temperatures of 20, 1000, and 1200°C using a pneumatically actuated system, a box furnace with SiC heating elements and a SiC fixture. Figure 7 shows a schematic of the fixture used for the tests. The shear test fixture, which was machined from sintered silicon carbide (Hexoloy-SA SiC), consists of two semicylindrical sections used to center the specimen, two pistons, one sleeve and an aligning disk. Two layers of ceramic paper were used to accommodate nonuniformities in the thickness of the specimen and dimensional changes that result from thermal expansion. To minimize Hertzian contacts and nonuniform loading, the parallelism between the ends of the specimen were maintained within 50 μm.

[2]Nippon Carbon Co., Japan

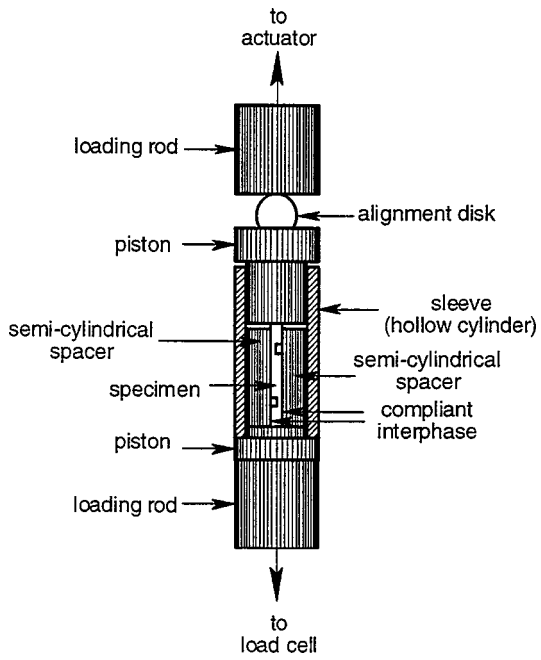


Fig. 7 Schematic of fixture used for the compression of double-notched specimens at elevated temperatures

For the tests at elevated temperatures, the specimens were heated to the test temperature at a rate of 60°C/min, followed by a soak period of 15 min to allow for thermal equilibrium. Temperature was monitored and controlled using a type ''S'' thermocouple which was positioned next to the fixture. The tests were conducted at a constant loading rate of 50 N/s.

Shear stress-rupture tests were also conducted at 1200°C for stresses of 35 and 17.5 MPa. The specimens were heated to 1200°C at a rate of 60°C/min followed by a soak period of 15 min. Then the specimens were mechanically loaded to the test stress at a rate of 50 N/s, and the load was maintained constant until the specimens failed and the time to failure was recorded.

Results and Discussion

Photomacrographs of two joined specimens with notches are shown in Figs. 8(*a*) and 8(*b*). Note that the composite material contains a large amount of porosity in the interlaminar regions,

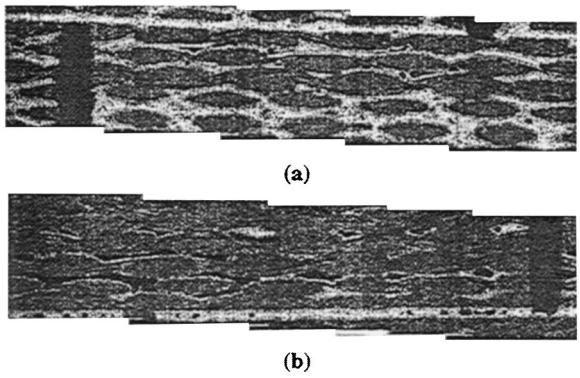


Fig. 8 (*a*) and (*b*) Photographs of double-notched joint specimens

Fig. 10 Optical micrograph of fracture surfaces of joined CVI SiC/SiC composites tested at 1200°C

which is characteristic of most composite materials synthesized by chemical vapor infiltration. It was also observed that the joint region contained porous regions [Fig. 8(*b*)].

Figure 9 shows typical curves of the average shear stress versus cross-head displacement obtained from the monotonic testing of specimens at 1000 and 1200°C. The average shear stress was calculated as follows:

$$\tau = \frac{P}{wh},$$

where **P** is the applied compressive load, **w** is the width of the specimen and **h** is the distance between the notches. Note that the curves in Fig. 9 are increasingly stiffer up to the peak load which suggests that there is no crack growth prior to failure. Also note that the compliance of the specimens (as determined from the instantaneous slope of the curves) increased with the test temperature.

Analysis of the specimens after the tests revealed that the specimens failed within the joint in a cohesive mode along a plane in between the notches. Micrographic examination of the fracture surfaces showed that some regions of the joint were not fully infiltrated, which was consistent with analysis of the edges of the specimen (Fig. 8). Figure 10 shows a optical micrograph of the fracture plane for a specimen tested at 1200°C.

Figure 11 summarizes the mechanical test results and indicates that the shear strength of the joints decreases with temperature from 92 MPa at ambient temperature to 71 MPa at 1200°C. The results from the shear stress-rupture tests are summarized in Fig. 12. It was found that the shear strength of the joints drops rapidly with time from 71 MPa, which is the ''virgin'' monotonic strength of the joints to 17.5 MPa after 14.3 h. These results are important because they revealed the weakness (i.e., rapid loss of strength with time) of a joint system that would have been considered robust based only on its virgin strength. Although additional work is still needed to identify the mechanisms responsible for the rapid loss of strength of these joints, it is well known that the formation and coalescence of cavities in siliconized silicon

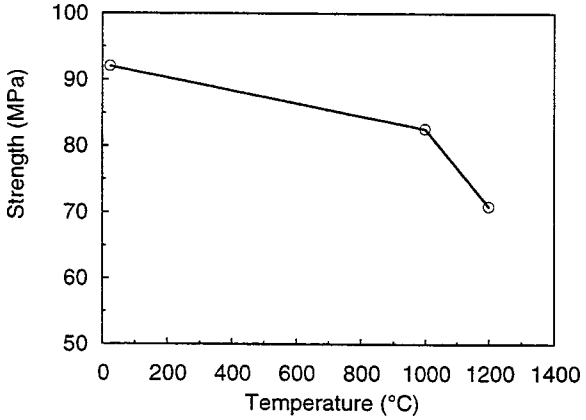


Fig. 11 Temperature dependence of apparent shear strength for joined SiC (CG-Nicalon™)/C/CVI-SiC composites

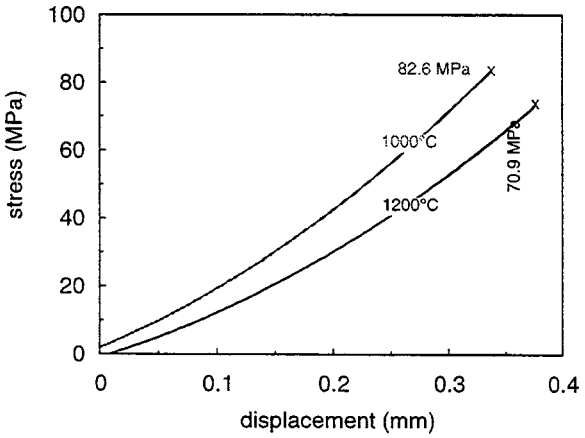


Fig. 9 Typical shear stress versus cross-head displacement curves obtained from the compression of double-notched joint specimens at high temperatures

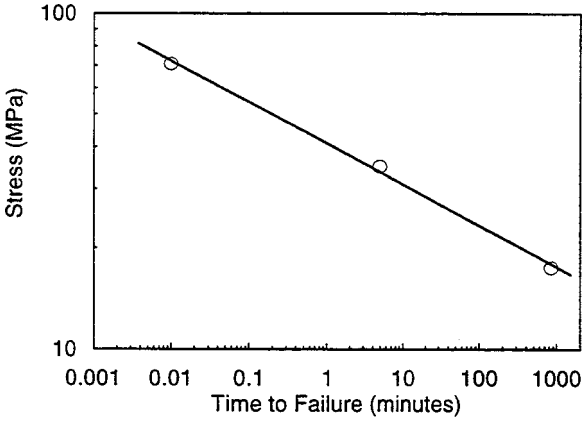


Fig. 12 Results from the shear stress-rupture evaluation of joints at 1200°C. The data point at the shortest time corresponds to the ''virgin'' shear strength of the joints.

carbide lead to the formation of cracks, and that cavity growth occurs very rapidly in siliconized silicon carbide ([16]). Therefore, the combination of high localized stresses around pores and notches and formation and coalescence of cavities are believed to be responsible for the rapid loss of strength exhibited by the joint material.

Conclusions

The shear strength and stress rupture behavior of joints of SiC (CG-Nicalon™) fiber-reinforced CVI-SiC matrix composites were determined by the compression of double-notched specimens. The joints were obtained using ARCJoinT and consisted of SiC plus small amounts of free silicon and titanium disilicide. From the evaluation of specimens with a notch separation of 4 mm, it was found that the apparent shear strength of the joints decreased from 92 MPa at room temperature to 71 MPa at 1200°C. However, it was found from stress-rupture testing in air at 1200°C that the strength of the joints decreased rapidly with time from a ''virgin'' monotonic strength of 71 MPa to 17.5 MPa after 14.3 h.

The conduct of extensive systematic tests of this nature (as a function of stress, time, temperature, and environment) is required to provide sufficient information for the generation of failure maps, and the kinetics of loss of strength. This information along with microstructural, chemical and fractographic examination will help identify the mechanisms responsible for the loss of strength and provide guidance for the optimization of the joints.

Acknowledgments

The authors would like to thank R. F. Dacek for his help in the preparation of specimens. The research activities carried out at ORNL were sponsored by the U.S. Department of Energy, Assistant Secretary for Energy Efficiency and Renewable Energy, Office of Industrial Technologies, Industrial Energy Efficiency Division and Continuous Fiber Ceramic Composites Program, under Contract No. DE-AC05-96OR22464 with Lockheed Martin Energy Research Corporation.

References

[1] Messler, Jr., R. W., 1993, ''Joining of Advanced Materials,'' Butterworth-Heinemann, Boston, MA.
[2] Schwartz, M. M., 1994, ''Joining of Composite Matrix Materials,'' ASM International, Materials Park, OH.
[3] Fragomeni J. M., and El-Rahaiby S. K., 1995, ''Review of Ceramic Joining Technology,'' Rept. No. 9, Ceramic Information Analysis Center, Purdue University, Indiana.
[4] Krenkel, W., Henke, T., and Mason, N., 1997, Key Eng. Mater., **127–131**, pp. 313–320.
[5] Singh, M., Kiser, J. D., and Farmer, S. C., 1997, Ceram. Eng. Sci. Proc., **18**, No. 3, pp. 161–166.
[6] Singh, M., and Kiser, J. D., 1997, 'Physics & Process Modeling and Propulsion R&T Conference,' NASA CP-10193 5:1–10.
[7] Singh, M., 1997, Scr. Mater., **34**, No. 8, pp. 1151–1154.
[8] Singh, M., 1997, Ind. Heat., **9**, pp. 91–93.
[9] Singh, M., 1998, J. Mater. Sci. Lett., **17**, No. 6, pp. 459–461.
[10] Singh, M., 1998, J. Mater. Sci., **33**, pp. 5781–5787.
[11] Singh, M., 1998, Adv. Mater. Proc., **10**, pp. 89–90.
[12] Singh, M., 1999, Key Eng. Mater., **164**, pp. 415–420.
[13] ASTM C 1292, ''Test method for shear strength of continuous fiber-reinforced advanced ceramics at ambient temperatures,'' American Society for Testing and Materials, Conshohocken, PA (1996).
[14] Lara-Curzio, E., and Ferber, M. K., ''Shear Strength of Continuous Fiber Reinforced Ceramic Composites,'' in *Thermal and Mechanical Test Methods and Behavior of Continuous Fiber Ceramic Composites*, ASTM STP 1309M. G. Jenkins, S. T. Gonczy, E. Lara-Curzio, N. E. Ashbaugh, and L. P. Zawada, Eds., American Society for Testing and Materials, Philadelphia, PA (1996).
[15] Singh, M., and Lara-Curzio, E., unpublished work (1999).
[16] Carroll, D. F., and Tressler, R. E., 1988, J. Am. Ceram. Soc., **71**, No. 6, pp. 472–477.

R. K. Ahluwalia
e-mail: walia@anl.gov

V. J. Novick
e-mail: vnovick@anl.gov

L. Zhang
e-mail: zhang@anl.gov

M. P. Sutaria
e-mail: manish sutaria@qmgate.anl.gov

J. P. Singh
e-mail: jp singh@qmgate.anl.gov

Argonne National Laboratory,
9700 South Cass Avenue,
Argonne, IL 60439-4841

Performance of a Vacuum Formed Chopped Ceramic Fiber Filter in a Reducing Environment

A laboratory-scale apparatus has been used for unattended, long duration, continuous flow through testing of a vacuum formed chopped ceramic fiber filter under reducing conditions at atmospheric pressure. Four candle specimens were exposed from 150 to 3550 h to 600°C gas containing 4 percent CO, 11 percent H_2*, 12 percent* CO_2*, 14 percent* H_2O*, 59 percent* N_2*, 1 ppmv NaCl, 50 ppmv* H_2S*, and 1000–2000 ppmw ash from a transport reactor operated in gasification mode. A database was established on pressure drop of the as-received and exposed filter as a function of face velocity and temperature. Tests were conducted to investigate the effects of back-pulse parameters on filter regenerability. Results are reported on the critical reservoir pressure and pulse duration for maintaining a stable saw-tooth profile of pressure drop across the filter element. Data are obtained to characterize the effect of chemical and thermal aging on the apparent bulk density of the filter, pore size distribution, fast fracture strength, and microstructure. It is suggested that the compliant filter undergoes a slow process of rigidization upon exposure to the test environment.* [DOI: 10.1115/1.1362669]

Introduction

Vacuum formed chopped ceramic fiber (VFCCF) filters made by Industrial Filter and Pump Manufacturing Company are being considered for high-temperature particulate removal applications in pressurized fluidized bed (PFBC) and integrated gasification combined cycle plants. Compared to isostatically pressed silicon carbide filters, VFCCF candles (trade name Fibrosic) have stated advantages of lightweight construction, stable ceramic oxide materials, high-temperature capability, shock and cracking resistance, and ease of surface treatment ([1]). At 0.27–0.34 g/cc apparent density, VFCCF candles are 70–80 percent lighter than their SiC counterparts. The fibers and binders are a blend of stable, pure ceramic oxides, primarily alumina and aluminosilicate materials rated for a maximum use temperature of 1400–1500°C. The differences in the thermochemical properties between the binder and fiber are negligible, as they are identical in composition. Due to the knitting tendency of the fibers during fabrication, VFCCF candle filters are stated to have an inherently high crack resistance. The random orientation of the fibers tends to blunt microcrack propagation. Finally, VFCCF candle filters can be treated with colloidal oxide materials to increase strength, add corrosion resistance, and control particle retention.

The purpose of this paper was to study the filtration behavior, evaluate regenerability, and assess the effects of extended thermal and chemical aging (more than 2000 h) on the microstructure and thermomechanical properties of the VFCCF filter. All tests were done under controlled reducing conditions in a laboratory setting. Previously, VFCCF candles have been tested under oxidizing conditions in a small PFBC facility at Argonne National Laboratory [2]. In two short-duration tests (exposure time <10 h) at 800°C, 9.2 bar and 5.1–10.2 cm/s face velocity, the candles demonstrated excellent resistance to thermal shock imposed by periodic back pulsing with N_2 gas. The measured collection efficiency was better than 99.97 percent for the 11-μm mean-diameter particles captured on the filter. Sections removed from the fibrosic candle were placed on a material surveillance coupon tree above the W_APF tubesheet at the AEP Tidd Demonstration Plant [3]. No microstructural changes were detected after 2815 h of static, flow-over exposure to the oxidizing PFBC environment. In separate simulated PFBC, high-temperature, oxidation corrosion tests with flow-through coupons, VFCCF material exhibited midbody fractures and hole formations under accelerated pulse cycling, tearing, and rupture when exposed to steam and air, and rigidization in presence of alkali and steam in air. VFCCF candle filters have also been exposed to a reducing gas from a pilot-scale transport reactor ([4]). The test conditions were 475°C temperature, 8.3 bar pressure, 6700-ppm particulate loading, and 1.4 cm/s filter face velocity. In 126-h exposure time, the filter was subjected to 70 back pulses of 0.25 s duration and 11.45 bar pressure. The particle size distribution had a median diameter of 11.3 μm upstream and 2.3 μm downstream of the filter. The candles were reported to have ruptured due to penetration of carbon that burnt *in situ* when the filter vessel was heated under oxidizing conditions. *In situ* combustion of carbon could raise the filter temperature to 800–900°C ([5]).

Experiment

Figure 1 presents the schematic of the experimental apparatus. A flanged mullite ceramic tube, of 3.75 in. diameter and 42 in. long, is used as the primary confinement. A second smaller ceramic tube 3.25 in. diameter and 30 in. long, is concentrically mounted inside the first tube to hold and position the test filter in the tube furnace. A stainless-steel flange, modified to mount the back pulse filter cleaning mechanism, provides a seal to the top of the outer ceramic tube. At the bottom of the outer tube, the flange of the inner tube and a stainless-steel 3 in. tee are sealed together with the bottom flange of the outer tube. The outer tube is supported from the bottom flange and is positioned inside a 6 in. diameter and 36 in. long, three-zone tube furnace.

The gases used to simulate the oxidizing or reducing environments are supplied to the branch of the tee below the furnace. All gas lines are heat traced to over 100°C, as are the two stainless-steel tees and the reducing union. Pressurized gas cylinders supply pure N_2 and CO_2 and their flow is controlled through rotameters. An electrolytic generator with an output capacity of 1.21 slpm supplies hydrogen. A pressurized gas cylinder supplies additional hydrogen, if necessary. The H_2 is regulated with mass flow controllers. Water vapor is generated by nebulizing water into drop-

Contributed by the Coals, Biomass, and Alternative Fuels Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the CBAF Division, July 1998; final revision received by the ASME Headquarters Oct. 2000. Editor: H. D. Nelson.

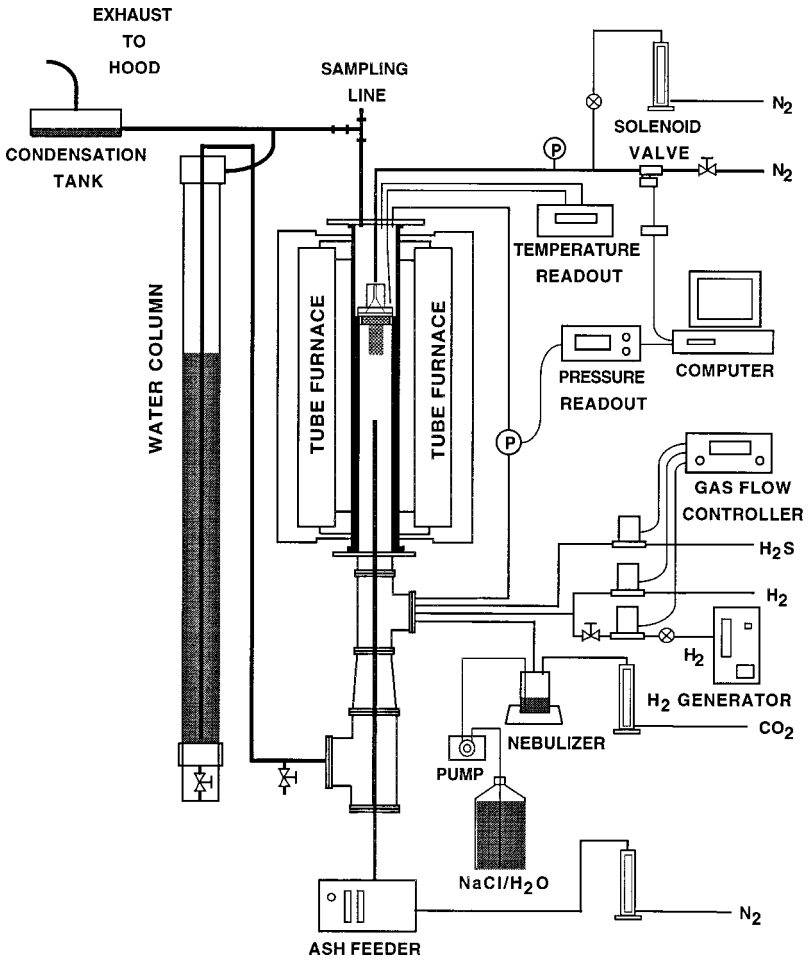


Fig. 1 Filter test apparatus

lets, which are carried by the CO_2. A peristaltic pump supplies a constant flow of distilled water to the nebulizer. The alkali challenge to the filter is in the form of sodium chloride, which is dissolved in water before nebulization. The flow rate of CO_2, the nebulizer power setting, the peristaltic pump setting, and the concentration of NaCl in the feed solution, control the amount of water vapor and alkali supplied to the experiment.

The N_2 gas added to the test environment is also used for dispersing ash into an aerosol and transporting it into the containment tube. The ash is regulated by a chain drive mechanism that transports it from a powder reservoir to the dispersal position using the adhesion of the ash onto the beads of the chain. The ash is blown off the bead by a jet of N_2 gas. The concentration of the aerosolized fly ash is determined by the speed of the chain supplying the ash and the volume flow of N_2 gas. The aerosolized fly ash is conveyed to the test filter through a long vertical 1/4 in. ceramic tube. The tube allows the carrier gas to maintain sufficient velocity to keep the fly ash suspended until the gas becomes hot enough so that, upon expansion in the 3 1/4 in. ceramic tube, it continues to carry the particles to the filter. Ash that is removed from the filter during the back-pulse process falls to the bottom of the apparatus. It cannot re-enter the ash supply tube because of the high velocity of gas in the 1/4 in. tube. Re-entrainment of the ash is minimized by allowing the back pulse of gas to flow out of the 3 in. tee below the furnace and through a water column, to the exhaust line. This keeps the removed ash entrained in the back-pulse gas until it reaches the cooler region outside the furnace where the upward flow is insufficient to suspend the fly ash. The water column also provides a safety mechanism to automatically release pressure from containment if it exceeds the height of the water column. This feature can protect equipment in a scenario where the filter blinds during unattended operation.

Pressure data is acquired every 2 s on a Mac IIx personal computer. A program was written to automatically acquire data, pulse the filter with back flow gas on a pressure limit or timed schedule, and to write 1 day's worth of data to a file for archiving. The ΔP across the filter is measured with a capacitance manometer, with an accuracy of better than 0.25 torr. The computer also controls the frequency, duration, and multiples of the back pulse of cold gas to the filter for cleaning purposes.

A schematic of the back pulse tube, filter holder, and filter geometry is given in Fig. 2. The back-pulse tube, with an outlet orifice of 1.2 mm, is positioned approximately 3 mm above the filter holder. The filter holder consists of two pieces, one is designed to distribute the back pulse gas over the entire cross section of the filter, while the other is designed to clamp the ceramic filter to the distributor. Both pieces are made of inconel for corrosion resistance. The distance between the back-pulse tube and the filter holder is critical to the operation and interpretation of the test results. If the tube is too close, a significant pressure drop can be created by the narrow gap and will be added to the overall ΔP measured by the capacitance manometer. If the tube is too far from the holder, a fraction of the back-pulse gas will not enter the distributor and not contribute to the filter's regeneration. The typical procedure was to find the minimum distance from the holder that did not contribute to an increase in ΔP as measured by the pressure transducer.

However, in some tests, a rapidly increasing ΔP that could not be recovered by back pulsing, was found to be related to the

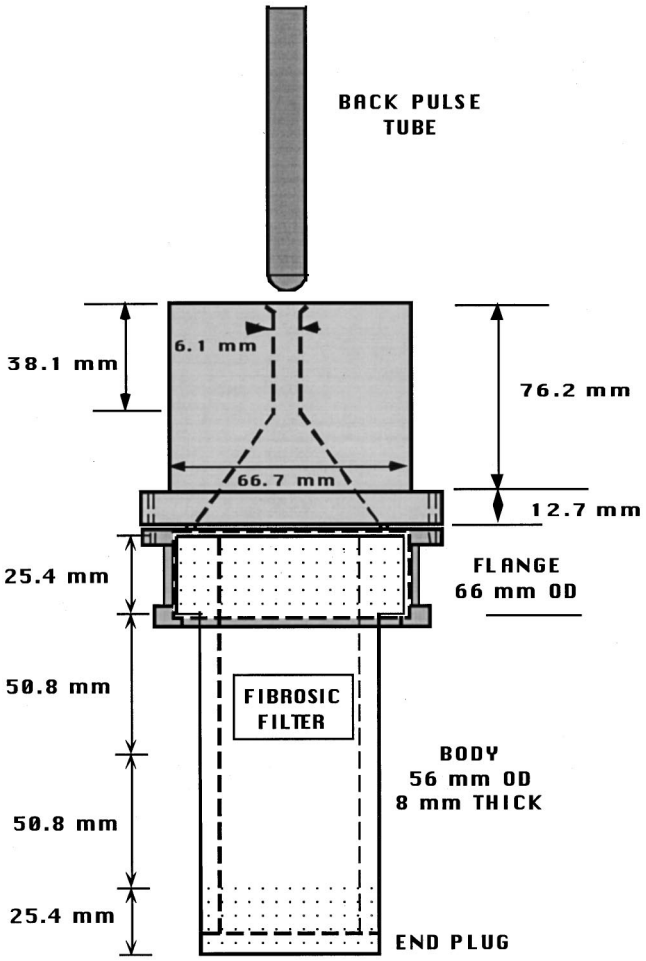


Fig. 2 Filter holder and back pulse system

positioning of the back-pulse tube and not to the blinding of the filter. Sometimes readjusting the position of the tube resolved the problem, other times the test had to be terminated. For the filter 3 test, the experiment was shut down, the apparatus was cleaned, and the test was restarted. It was determined that a small flow of N_2 (nominally 0.51 pm) at low pressure, continuously admitted through the back-pulse tube, eliminated the phenomenon.

The gas composition used in the tests was intended to be representative of a gasifier environment at atmospheric pressure. It was assumed that a desulfurization unit upstream of the filter could reduce H_2S concentration to less than 200 ppmv. Table 1 lists the target composition of the feed. For safety reasons, the input gas did not contain CO. Instead, the water-gas shift reaction was relied upon to produce CO through reduction of CO_2 in the chamber. During the test campaign, significant condensation of water vapor was noted in the cooler regions of the ash collection tee at the bottom of the assembly. To prevent unseen problems later in the campaign, the H_2O feed rate was reduced below the target value. Some difficulty was experienced early on in regenerating the fibrosic filter. Late in the campaign, it was realized that the difficulty was at least partly related to the incorrect positioning of the back-pulse tube relative to the venturi. To enhance the probability of running a successful long-duration test, the H_2S concentration too was lowered below the target value. Based on the experience gained from the tests, it is believed that both steps were unnecessary. Note that the quantity of ash challenging the filter was limited by the beaded chain supply mechanism of the ash feeder.

The ash used to challenge the filters came from the transport reactor demonstration unit run number P047. Table 2 lists the composition of the ash analyzed using x-ray fluorescence and x-ray diffraction. The ash was stored in a laboratory furnace at about 110°C for at least 1 day before placing it in the generator for dispersal. The intent was to improve aerosolization by removing moisture that might cause the ash to adhere strongly and impede efficient dispersal. The particle size distribution was measured with an eight-stage cascade impactor having cut points from 0.3 to 10 μm. The measured size distribution was graphed on a

Table 1 Composition of gas used in exposure tests. The equilibrium column is calculated from feed composition and gas phase chemistry.

	Target	Target Equilibrium	Filter 1 (Avg. Feed)	Filter 2 (Avg. Feed)	Filter 3 (Avg. Feed)	Filter 4 (Avg. Feed)
CO, %	0	4.3	0	0	0	0
H_2, %	15	10.7	15.1	13.8	13.5	15.3
$C0_2$, %	16	11.7	16.1	15.4	14.8	14.3
N_2, %	59	59	60	65.4	66.1	64.8
H_2O, %	10	14.3	8.8	5.4	5.6	5.6
H_2S, ppmv	200	200	75	21	31	32
NaCl, ppm	1	1	0	1.14	0.891	1.38
Ash, ppmw	2000	2000	1265	1847	1285	1625
Exposure Time, h			148	307	3532	263.3
No. of Cleaning Pulses			535	612	8353	1165
Temperature, °C	600	600	600	600	600	588
Pressure, atm	1	1	1	1	1	1

Table 2 TRDU ash analysis

Oxides	Weight %	Elemental	Weight %
$Si0_2$	50.4	Si	46.6
$A1_20_3$	20.1	Al	21.7
Fe_20_3	1 .4	Fe	2.0
$Ti0_2$	0.7	Ti	0.9
P_20_5	0.8	P	0.7
Ca0	12.8	Ca	18.1
Mg0	6.9	Mg	8.3
Na_20	0.7	Na	1.1
K_20	0	K	0
$S0_3$	1.1	S	0.8

log-probability plot and the mass median aerodynamic diameter (MMAD) of the dispersed ash was extrapolated to be 15.6 μm.

Figure 2 also presents the schematic of the fibrosic filters tested. The mean fiber diameter is specified as 3 μm and the mean binder particle diameter as 20 nm. The supplier impregnates the clean and dirty sides of the filter element with alumina to ''normalize'' the pore size and to toughen the skin. A filter specimen is 152-mm long, 8-mm thick, and has nominal outer diameter (OD) of 56 mm. It has a flange, 66 mm in OD and 25.4 mm in length. The flange is treated post fabrication to triple its density and increase its strength. Portions of the filter bottom (25.4 mm) and the body (50.8 mm) adjacent to the flange are rendered impervious by the same densification process. Thus, only 50.8 mm of the body is permeable to flow giving a net filtration area of 89.37 cm^2. The test specimen was sealed to the filter holder with a flexible ceramic fiber felt mat, sandwiched between the filter and the collar and diverter holder assemblies. The assembled filter holder was then sealed to the inner ceramic tube with more of the ceramic fiber felt mat, using the weight of the filter holder assembly to provide sufficient compression of the mat. This sealing mechanism was shown to be adequate, based on a comparison of ΔP measurements as a function of flow rate between a neoprene rubber sealing fixture and the mechanism described previously. To within the limits of the capacitance manometer, the ΔP measurements at each flow rate were identical with both sealing techniques, implying that there is no significant leakage around the ceramic mat seal.

Filter Performance

Pressure Drop Behavior. A parametric study was conducted on filter 4 to determine the relationship between ΔP, gas flow rate, temperature, and the condition of the filter. Figure 3 presents the data from a series of experiments in which ΔP was measured as a function of N_2 flow rate and furnace temperature. The open symbols represent the clean filter data. The pressure drop is observed to increase with gas temperature because of higher face velocity and gas viscosity. The filled symbols represent the data from the exposed filter. The variability in pressure drop data at 600°C (inverted filled triangles) for a given flow rate reflects the effects of repeated back pulses of N_2 gas, pulse duration, reservoir pressure, and the time between pulses. The data for 500°C (filled triangles) is representative of the day-to-day variation in pressure drop as a function of flow rate. The difference between ΔP of the exposed and clean filter for identical temperature and flow rate is a measure of the additional resistance to gas flow offered by the layer of adherent ash that builds up over numerous cleaning cycles.

The pressure drop data can be reduced by calculating the effective permeability or permeance (k_{eff}) from the Darcy-Forchheimer equation.

$$k_{\text{eff}} = \frac{\mu v (1 + \text{Re})}{\Delta P}$$

$$\text{Re} = \frac{\beta \rho v}{\mu} \tag{1}$$

where ρ is the gas density, μ is the gas viscosity, v is the face velocity, and β is an empirical coefficient to account for inertial effects. With $\beta = 0$ and data for all combinations of temperature and flow rate included, an average k_{eff} of 4.05 ± 0.63 Å is inferred for clean filter 4. The error bar is somewhat larger for filter 1 because the data for 1 slpm was included. The pressure drop at this flow rate was always inconsistent. Disregarding 1 slpm data, the average k_{eff} for filter 1 becomes 4.17 ± 0.58 Å. Figure 4 plots the effective permeability as a function of N_2 gas temperature and velocity. There is a visible trend of k_{eff} increasing with gas temperature and face velocity. At a given temperature, the variability in k_{eff} is due to different face velocities. Similarly, the variability in k_{eff} at a constant face velocity is due to differences in gas temperature. The pressure drop data can be reduced better by including the inertial effects in Eq. (1). The filter to filter variation in k_{eff}, however, is even larger even though they came from the same manufactured lot. Note that the filtration area was calculated and assumed to be the same for different specimens. However, the potential difference in filtration area is not expected to account for the observed variation in k_{eff} between the filters.

The pressure drops for filters 2 and 4 were measured as a function of N_2 gas flow rate and temperature during the cool-down phase of the exposure tests summarized in Table 1. The average effective permeability for filter 2 at 20°C is 2.88 ± 0.39 Å. For filter 4, the average k_{eff} over the temperature range 20–600°C is

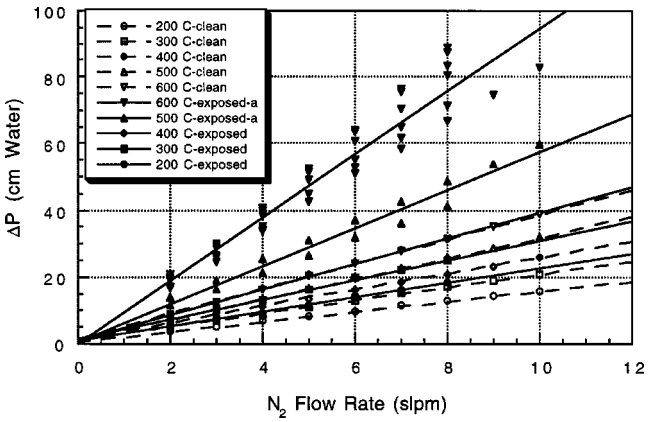


Fig. 3 Pressure drop behavior of filter 4 as a function of gas flow rate and temperature, before and after exposure

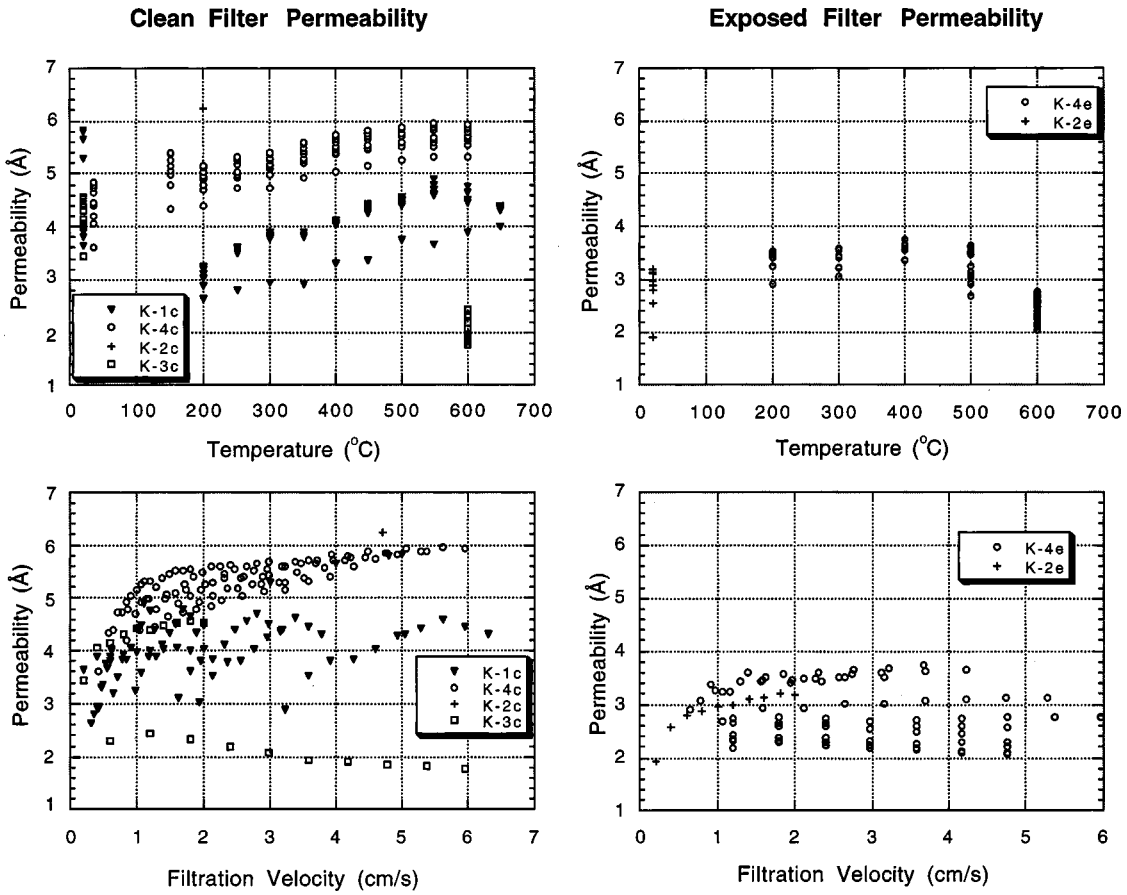


Fig. 4 Derived permeability of VFCCF filters as a function of face velocity and temperature, before (symbol c in legend) and after exposure (symbol e in legend)

2.86 ± 0.52 Å. One effect of exposure to ash is to reduce the filter-to-filter variation in k_{eff} measured with the as-received specimens. On exposure, the effective permeability reduces because of the ash cake that forms on the outer surface and possible penetration of ash into the interstices of the filter. The contribution of ash to k_{eff} can be deduced by representing it as two resistances in series

$$\frac{1}{k_{\text{eff}}} = \frac{1}{k_{\text{eff}}^{0}} + \frac{1}{k_{\text{eff}}^{\text{ash}}} \tag{2}$$

The data suggest that for filters from the same manufactured lot challenged by the same ash at identical operating conditions, $k_{\text{eff}}^{\text{ash}}$ is apparently inversely proportional to the effective permeability of the virgin specimen k_{eff}^{0}. For filter 4, $k_{\text{eff}}^{\text{ash}}$ is estimated as 6.31 ± 0.96 Å. Post-test examination of one of the exposed filters revealed nearly uniform coating of the residual ash and a cake thickness of 0.51 mm.

Figure 4 also depicts the dependence of k_{eff} for the exposed filters on gas temperature and face velocity. The effective permeability appears to increase with temperature over the range 20–400°C. Whereas k_{eff} for the clean filter becomes flat above 500°C or slightly decreases, it clearly decreases with temperature above 400°C for the exposed filters. Also, the effective permeability for the exposed filters is relatively constant for face velocity exceeding 1 cm/s.

Filter Collection Efficiency. The particle collection efficiency of the fibrosic filter was measured at 1435 h into the 3530-h exposure test of filter 3. A cascade impactor was installed in the exhaust line directly above the outlet of the filter holder assembly. The amount collected by the stages of the cascade impactor was compared to mass distribution of the fly ash as dispersed in separate offline tests of the ash feeder. In addition, the total mass output per unit time of the ash feeder was measured. The average MMAD of the particulates aerosolized by the ash feeder was 15.6 μm. The average output of the ash feeder was 8.15 mg/min with an aerosolization efficiency of 45.8 percent. A total of 0.29 mg of ash was collected after 6 h of sampling at 1 slpm downstream of the filter. Analysis of the cascade impactor and ash feeder data show that 99.96 percent of particles larger than 16 μm, 99.83 percent of particles between 7.85 μm and 16 μm, 99.86 percent of particles between 4.24 μm and 7.85 μm, 99.89 percent of particles between 2.05 μm and 4.25 μm, and 99.92 percent of particles between 1.09 μm and 2.05 μm were collected by the filter. The total collection efficiency, based on the mass collected downstream of the filter and the estimated mass challenging the filter, was determined to be 99.92 percent.

Filter Loading and Cleaning Behavior. A test sequence was designed to explore if any one constituent of the environment caused an unreasonable increase in ΔP across filter 1 or problems with filter regeneration. Starting with pure N_2 and a furnace temperature of 200°C, each constituent was added and the filter behavior was followed for up to 24 h. With pure N_2 and 200°C furnace temperature, a pressure drop of 31 cm H_2O was noted at 4 cm/s face velocity. Ash was added and the cleaning schedule was set as a double pulse of 200 ms duration and 0.5 s apart, every 20 min from a N_2 reservoir at 9.5 bar. Over 4 h, a saw-tooth profile developed with baseline ΔP gradually increasing and stabilizing at 58 cm H_2O. A series of 10 pulses over a 6-min period from a N_2 tank at 11.2 bar lowered the baseline ΔP to 51-cm H_2O. Additional 15 pulses over the same period further decreased the baseline ΔP to 48-cm H_2O. Over the next 4 h, it was demonstrated

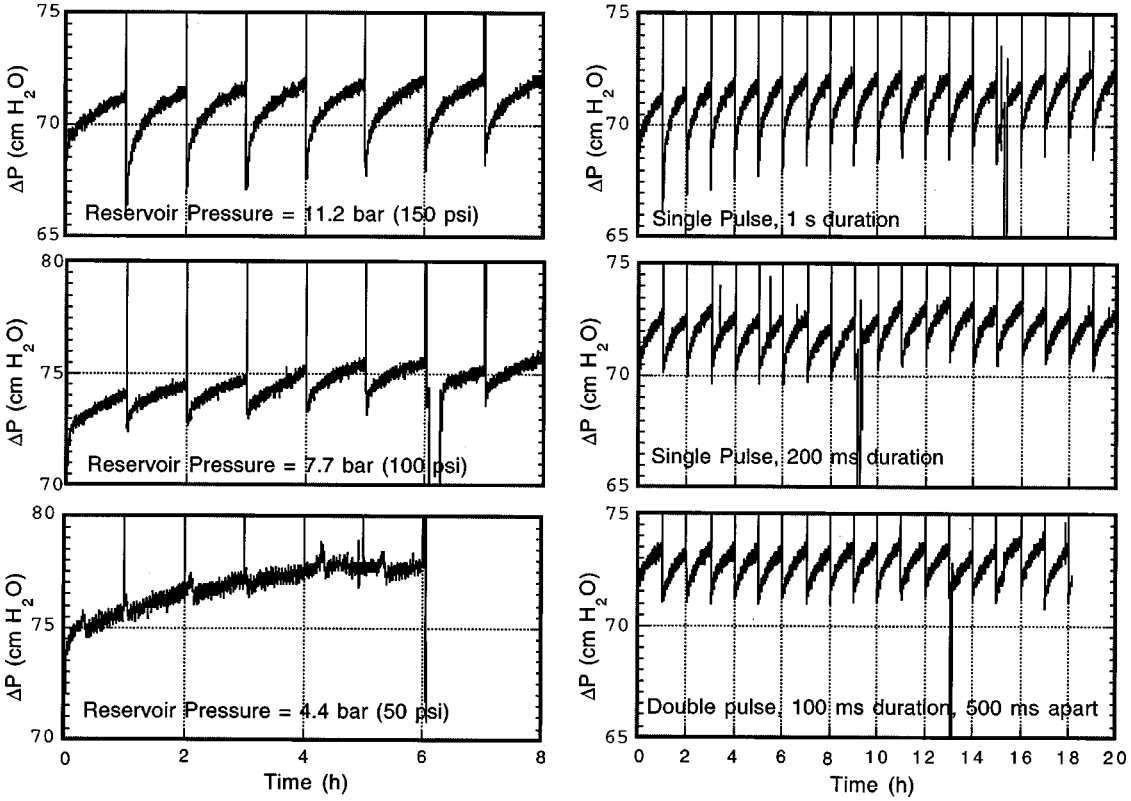


Fig. 5 Measured ΔP across the filter as a function of time at 600°C. The graphs show the effective cleanability of the filter as a function of reservoir pressure for a 1-s single pulse, and as a function of pulse duration at 11.2 bar reservoir pressure

that a stable saw-tooth ΔP profile could be maintained with a single pulse of 0.5 or 1 s duration and a reservoir pressure of 11.2 bar. The stable baseline ΔP was 51–52 cm H_2O and the increment in ΔP was 2 cm H_2O between consecutive pulses with 20-min interval.

The furnace temperature was raised in increments of 100°C, N_2 flow was adjusted to maintain relatively constant filtration velocity, and ΔP was monitored for 24 h at each temperature setting. No deviations from the stable loading and regeneration cycle were noted. The additions of CO_2, H_2, and H_2O did not affect the cycle, nor did the addition of alkali. Admission of H_2S appeared to increase the pressure drop. On turning off H_2S flow, it took about 24 h for the baseline ΔP to reach the original level prior to addition of H_2S.

Tests with filter 4 were dedicated to investigating the effects of back pulse parameters on the effectiveness of filter regeneration at 500 and 600°C. Figure 5 presents the data on the effect of reservoir pressure on ΔP profile. At 600°C furnace temperature, 1-s pulse duration, 60-min pulse interval, and a stable repetitive saw-tooth profile is established at 11.2 bar reservoir pressure. Under the same conditions, a saw-tooth profile is also evident at 7.8 bar reservoir pressure but it is nonrepetitive because of the creep in baseline ΔP. There is only a partial recovery in ΔP with each back pulse so that the pressure drop rises from one pulse cycle to the next. A saw-tooth profile is not observed at 4.4 bar reservoir pressure. At this operating condition, the back pulse is ineffective in regenerating the filter and there is little or no recovery in pressure drop. Thus, the three reservoir pressure levels can be associated with the limiting conditions of complete, partial, and zero recovery of ΔP with back pulsing.

Figure 5 also compares data on filter regenerability as a function of pulse duration and frequency at 600°C furnace temperature and 11.2 bar reservoir pressure. The data shows that a pulse duration of 200 ms is sufficient to repetitively regenerate the filter. The baseline pressure is somewhat lower with a 1-s pulse duration but the gas consumption is five times as large. A stable saw-tooth profile could also be established with a double pulse, 100 ms in duration, and 500 ms apart. However, a double pulse does not appear to offer any advantages over a single pulse of the same total duration.

Tests were also run to examine the effect of pulse interval on filter regenerability. At 600°C, 11.2 bar reservoir pressure and 1-s pulse duration, the differences in ΔP traces between 20, 60, and 120 min pulse interval tests were considered insignificant. In each case, ΔP profiles were repetitive, the baseline pressures were similar, and the increase in pressure drop between pulses was proportional to the time interval.

The above results suggest that there is a critical reservoir pressure below which the VFCCF candle cannot be regenerated by the simple back-pulse technique. Above this critical pressure, a short pulse is as effective in maintaining a stable saw-tooth profile as a long pulse although the baseline pressure is affected by pulse duration. Pulse schedule, single or multiple pulses, and interval between pulses, does not appear to significantly impact filter regenerability. For the solenoid valve used in the test apparatus, a 18-ms opening and 30-ms closing time, 200-ms pulse duration was adequate to regenerate the filter. At 11.2 bar reservoir pressure, the required specific flow rate of the pulse gas is 1.35 slpm/cm^2 of filtration area. Thus, the projected compressed gas consumption for a standard 1500-mm long, 60-mm OD candle is 12.7 sl/pulse/filter at 200-ms pulse duration.

Mechanical Behavior

The four fibrosic filters exposed to the reducing conditions were examined for changes in physical properties, mechanical properties, and microstructure. Two virgin candle filters were obtained to establish the reference properties. These were identical to the

Table 3 Summary of physical and mechanical properties of virgin (filter 0) and exposed fibrosic filter specimens

	Filter 0	Filter 1	Filter 4	Filter 2	Filter 3
Exposure Time, h	0	148	243	307	3532
Bulk Density, g/cm^3 (Geometrical Method)	0.293	0.325	0.413	0.370	0.387
Bulk Density, g/cm^3 (Mercury Intrusion)	0.341	0.400		0.350	0.420
Porosity	0.85 (0.73)	0.68 (0.59)		0.77 (0.66)	0.68 (0.59)
Median Pore Size, μm	40.31	24.31		24.17	23.51
Cumulative Intrusion Volume, cc/g	2.5	2.0		2.25	2.0
Fracture Strength, kPa (O-Ring Compression)	401.30 ± 23.00	843.11 ± 30.00	1533.65 ± 124.75	1741.60 ± 10.75	1819.25 ± 186.15
Flexural Strength, kPa (Three-Point Bending)	515.95 ± 129.98				
Elastic Modulus, kPa	43 ± 21				

filters that were exposed in that they came from the same production lot number. The virgin filters, however, did not have a triple density flange or the end plugs at the opposing ends. They had the same inner diameter, 39.4 mm, but somewhat larger outer diameter, 60 mm. The filters exposed to the tests were actually machined down to 55.4-mm OD from 60-mm OD in order to fit the experimental apparatus.

Physical Properties. The apparent density of the filter material was measured by a geometrical method. Rectangular specimens were cut by a diamond saw and ground to remove surface roughness. The apparent bulk density, calculated from the measured mass and volume of the rectangular specimens, is listed in Table 3. A bulk density of 0.293 g/cm^3 was measured for the virgin filter. The bulk density was observed to increase with exposure; the average density of the four exposed filter specimens was 0.374 ± 0.037 g/cm^3. The phenomenon of increase in apparent bulk density with exposure to ash, steam, alkali, and H_2S, in a reducing environment, was confirmed with data from mercury intrusion porosimetry.

The porosity of the filter materials was measured by mercury intrusion porosimetry. This method also provided bulk density, total open porosity, and the pore size distribution. The bulk density measured with the mercury intrusion porosimetry was 0.341 g/cm^3 for the virgin filter and 0.390 ± 0.036 g/cm^3 for the exposed filters. As with the geometrical method, the bulk density increased with the time of exposure.

Figure 6 presents the measured cumulative intrusion volume as a function of pore diameter for the as-received and exposed filters. The virgin filter has a bimodal size distribution with about 15 percent of the intrusion volume associated with the submicron pores. Pores in the 10–90-μm diameter range account for more than 80 percent of the intrusion volume. The median pore diameter is 40.3 μm and the cumulative intrusion volume is 2.5 cc/g.

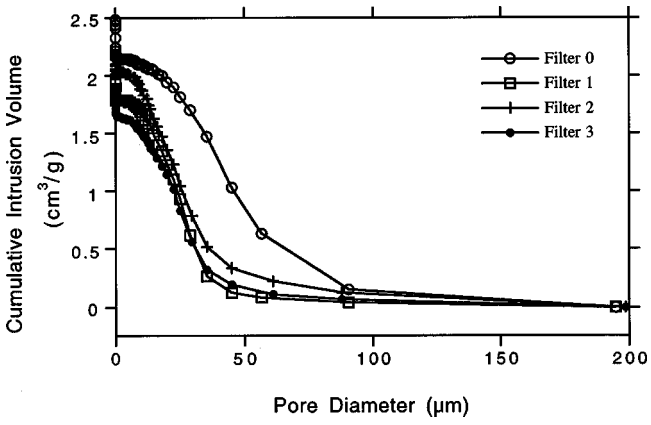


Fig. 6 Effect of exposure on pore size distribution represented as cumulative intrusion volume less than the stated pore diameter

The virgin filter is estimated to have a porosity of 85 percent. Upon exposure, the cumulative intrusion volume decreased to 2.12 ± 0.12 cc/g as did the median pore size to 23.91 ± 0.4 μm. The data suggest that the intrusion volume associated with the submicron pores experienced only minor change. Much of the decrease in intrusion volume is due to partial closing of the macropores. It appears that the intrusion volume and the median pore size decrease precipitously (within 150 h) as the filter is exposed to the ash-laden flow. Subsequently, the rate of decrease of intrusion volume median pore volume appears to slow down.

Also listed in Table 3 is the filter porosity estimated from the cumulative intrusion volume and the bulk density of the virgin filter. Two estimates are given, one is based on the bulk density deduced from the geometrical method and the other from the mercury intrusion porosimetry. The data indicate that the filter porosity decreases by 14.7 ± 5.3 percent due to exposure to ash.

Mechanical Properties

O-Ring Compression Tests. O-ring specimens, 0.5-in. (1.27-cm) wide, were cut from each candle filter using a diamond saw. Two specimens were tested for each exposed filter and four for the virgin filter. These specimens were loaded to failure in a diametral compression mode on an Instron mechanical testing machine at a crosshead speed of 0.15 cm/min. The maximum stress is developed at the inner diameter across the load points. The fast fracture strength σ_f, is given by

$$\sigma_f = \frac{PK}{\pi b l} \tag{3}$$

where P is the fracture load, K is a function of the ratio of inner to outer diameters, b is the outer radius of the specimen, and l is the length of the ring. The calculated fracture strength is listed in Table 3.

A typical load-displacement plot of a filter tested in the O-ring compression test is presented in Fig. 7. The plot shows a characteristic composite material behavior. All specimens failed by debonding of the ceramic layers within the filter walls.

The fast fracture strength of the virgin filter was measured to be 401.3 ± 23 kPa. It increased monotonically with time of exposure to the reducing gas containing alkali, H_2S, water vapor, and ash. The fracture strength increased to 843.1 ± 30 kPa after 148 h, 1741.6 ± 10.8 kPa after 307 h and 1819.2 ± 186.2 kPa after 3532 h. The limited data suggests that the chopped ceramic fiber filter experiences a rapid improvement in strength for the first 300 h of exposure and a gradual increase thereafter. No degradation in strength was observed within 3500 h of exposure.

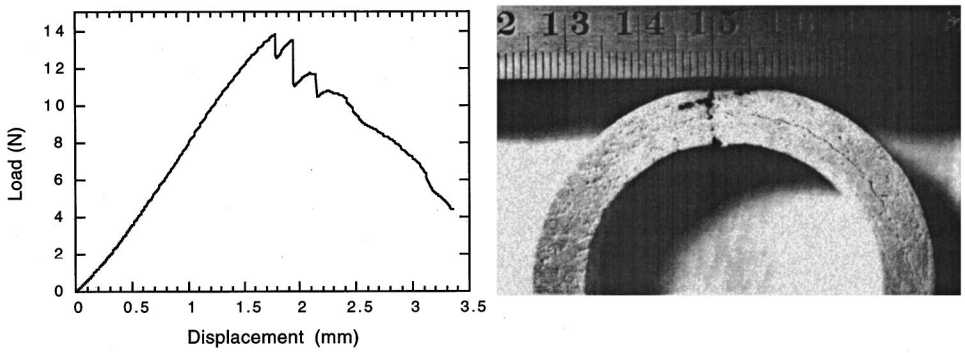


Fig. 7 Typical load-displacement plot and failure mode in O-ring compression tests

Besides fracture stress, there are other indicators of the compliant filter becoming rigidized upon exposure to the reducing gas. In the O-ring compression test, the uncompensated deflection of the specimens at peak load was observed to generally decrease with time of exposure. The measured deflection was 1.49 ± 0.29 mm after 148 h, 0.64 ± 0.03 mm after 243 h, 1.05 ± 0.18 mm after 307 h, and 0.66 ± 0.01 mm after 3589 h. Similarly, the slope of the load-deflection curve, which should be proportional to the elastic modulus, was 16.7 N/mm for the ash-received filter specimen, 9.52 ± 1.32 N/mm after 148 h, 46 ± 0.8 N/mm after 243 h, 27.61 ± 9.81 N/mm after 307 h, and 50.15 ± 4.65 N/mm after 3532 h. More data is required to confirm the increase in stiffness with time of exposure.

Three-Point Bending Tests. Flexural strength was measured on rectangular bar specimens in a three-point bending mode. The specimens were cut by a diamond saw and polished on a 600 grade SiC grinding paper. The final dimensions were 48.25 mm $\times 8.90$ mm $\times 5.10$ mm. A loading span of 38 mm and a crosshead speed of 0.15 cm/min were used for strength measurement. Figure 8 shows a typical load-displacement plot for a filter specimen. The plot depicts a graceful failure rather than a brittle failure. The flexural strength and the elastic modulus were calculated using the equations

$$\sigma_f = \frac{3PL}{2bh^2}$$

$$E = \frac{PL^3}{12bh^3 y_{\max}} \tag{4}$$

where P is the maximum load, L is the loading span, b is the width of the specimen, h is the height of the specimen, and $y_{\max}$ is the deflection. The average flexural strength for the seven replicate virgin specimens is calculated to be 515.95 ± 129.98 kPa; the average elastic modulus is 43 ± 21 kPa.

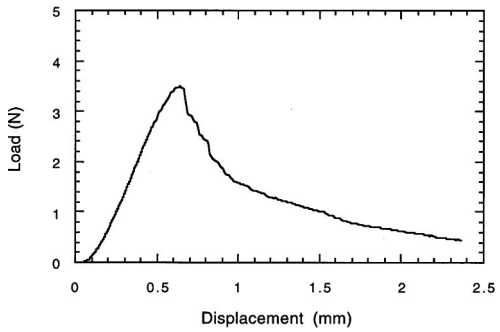


Fig. 8 Load-displacement plot from three-point bending tests

Microstructure. Scanning electron microscopy (SEM), coupled with energy-dispersive x-ray (EDX) analysis, was performed on the fractured surfaces (radial cross section) of filter specimens to investigate penetration of ash particulates and/or chemical reactions. Figure 9 shows a SEM micrograph of the virgin filter. It indicates a composite material with a high volume of fiber content. EDX analysis revealed alumino silicate peaks for both the fiber and the matrix.

SEM and EDX analyses were performed on the inner and outer surfaces of the cylinder to investigate ash penetration. Figure 10 shows a large amount of debris on both the inside and outside layers of filter 1, indicating significant ash penetration. During the experiment, this filter proved difficult to regenerate by back pulsing. EDX analysis of both fibers and matrix did not show any difference between the inner and outer surfaces. Furthermore, the EDX peaks for this filter were the same as those for the virgin filter. Thus, there was no significant chemical attack during exposure. For filter 2, there was little evidence of ash particles accumulating on the inside or outside surfaces. For filter 3, ash particles are visible on the exposed outside surface but not on the inside surface. This is indicative of little ash penetration and an effective regeneration cycle. For filter 4, ash accumulation on the outside surface is evident with some indication of ash breakthrough to the inner surface.

Summary and Conclusions

A laboratory-scale apparatus has been assembled for unattended, long duration, continuous, flow-through testing of advanced, high-temperature, particulate filters for applications in pressurized fluidized bed combustion and integrated gasification combined cycle plants. It includes a provision for on-line, computer-controlled, back-pulse cleaning of the fouled filter element. The apparatus has been used to test the VFCCF filter specimens supplied by the Industrial Filters and Pump Manufacturing Company. The objectives of the tests were to study the pressure drop behavior, evaluate regenerability, and assess the effects of thermal and chemical aging on the microstructure and thermomechanical properties of the VFCCF filter. Four specimens were exposed from 150 to 355 h to a 600°C reducing gas environment containing 4 percent CO, 11 percent H_2, 12 percent CO_2, 14 percent H_2O, 59 percent N_2, 1 ppmv NaCl, 50 ppmv H_2S, and 1000–2000 ppmw ash. The ash used to challenge the filters came from the Transport Reactor Demonstration Unit at the University of North Dakota Energy and Environment Research Center. Some important results from the test campaign are summarized below:

• A database was established on pressure drop of as-received filters as a function of face velocity and temperature. The observed pressure drop can be correlated with an effective permeability of 4.2 ± 0.6 Å. The data showed a visible trend of effective permeability increasing with gas temperature and face velocity.

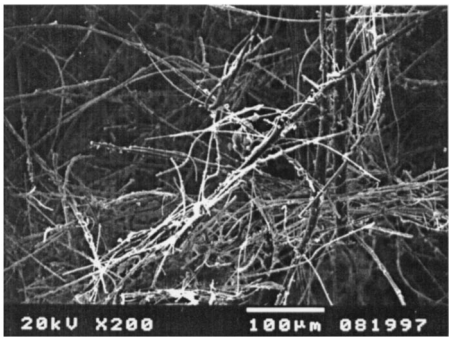


Fig. 9 SEM micrograph of the virgin filter showing composite material with a high fiber content

There was considerable filter-to-filter variation in effective permeability even though the specimens were from the same manufactured lot.

• On exposure, the effective permeability decreased to 2.9 ± 0.5 Å over the temperature range 20–600°C. One effect of exposure to ash is to reduce the filter-to-filter variation in effective permeability measured with the as-received specimens. Post-test examination of the exposed filters revealed nearly uniform coating of the residual ash and a cake thickness of 0.5–1 mm.

• The mean collection efficiency of the filler, as measured with a six-stage cascade impactor, is 99.92 percent for the challenge ash with a mass median aerodynamic diameter of 15.6 μm.

• A number of tests were conducted to investigate the effects of back-pulse parameters on filter regenerability. At 600°C, a stable saw-tooth profile was established with a single 1-s pulse from a reservoir at 11.2 bar (150 psig). A saw-tooth profile was also observed at 7.8 bar (100 psig) reservoir pressure although the baseline pressure kept increasing slowly with time. A saw-tooth profile could not be established at 4.4 bar (50 psig) reservoir pressure. At 11.2 bar, a single pulse of 200 ms duration was also adequate in repetitively regenerating the filter to a stable baseline pressure. A double pulse of 100 ms duration and 500 ms apart, stabilized the filter at a slightly higher baseline ΔP.

• The apparent bulk density of the filter, as measured by a geometrical method, was 0.293 g/cc for the virgin specimen and 0.387 g/cc after 3550 h of exposure. This trend of increase in bulk density with time of exposure was confirmed by mercury intrusion porosimetry data.

• The virgin filter has bimodal pore size distribution with 15 percent intrusion volume associated with submicron pores and >80 percent accounted for by pores in the 10–90-μm diameter range. Upon exposure, the cumulative intrusion volume decreased from 2.5 to 2.12 ± 0.12 cc/g and the median pore diameter from 40.3 to 23.9 ± 5.3 μm.

• The fast fracture strength of the filter specimens was mea-

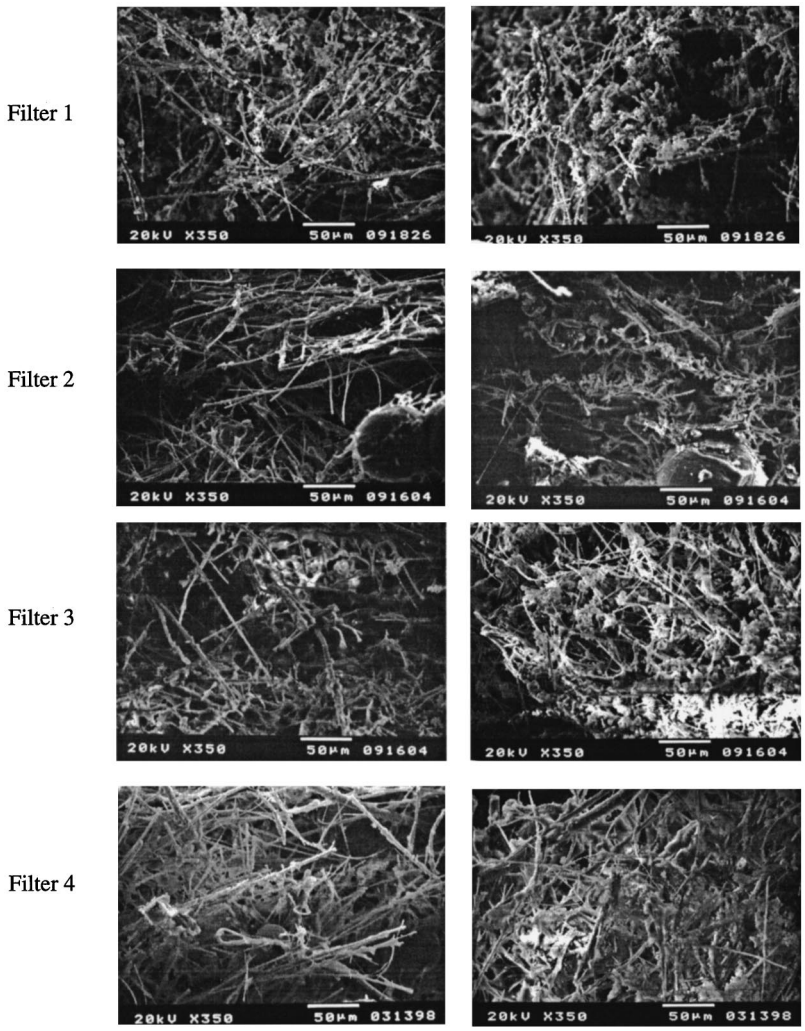


Fig. 10 SEM micrographs of the inner and outer surfaces of the exposed filters

sured with O-ring compression tests. The load-displacement plots exhibit a composite material behavior. All specimens failed by debonding of the ceramic layers within the filter walls. The fast fracture strength increased from 401.3 ± 23 kPa for the virgin filter to 843.1 ± 30 kPa after 148 h 1741.6 ± 10.8 kPa after 307 h, and 1819.2 ± 186.2 kPa after 3532 h.

• Based on the measured changes in bulk density, cumulative intrusion volume, median pore diameter, fast fracture strength, deflection at peak load, and the slope of the load-deflection curve, it is suggested that the compliant VFCCF filter becomes rigidized upon exposure to the reducing environment containing H_2O vapor, trace amounts of NaCl and H_2S, and fly ash.

• SEM and EDX analyses of the virgin filter indicate a composite material with a high fiber content. There is some knitting of fibers, somewhat random orientation of fibers, and the fiber diameter is <10 μm. One of the filter specimens showed a large amount of debris on both the inside and outside layers. During the experiment, this filter had proved difficult to regenerate. The EDX peaks were the same as those for the virgin filters indicating that there was no significant chemical attack during exposure. For the other filter specimens, ash particles were visible on the exposed outside surface but not on the inside surface.

Acknowledgments

Research was sponsored by the U.S. Department of Energy's Federal Energy Technology Center. Dr. Norman Holcombe was the Contracting Officer's Representative. We thank Mr. Paul Eggerstedt of Industrial Filter and Pump Manufacturing Company for supplying the filter specimens and Dr. John Hurley of University of North Dakota for supplying the ash.

References

[1] Eggerstedt, P. M., 1995, ''Lightweight Ceramic Filter Components: Evaluation and Application,'' *Proceedings of the Advanced Coal-Fired Power System' 95 Review Meeting*, Morgantown, WV, Vol. 1, pp. 140–149.

[2] Lee, S. H. D., Eggerstedt, P., and Zievers, J. F., 1996, ''Preliminary Evaluation of the Fibrosic Candle Filter for Particulate Control in PFBC,'' J. Inst. Energy, **9**, pp. 87–93.

[3] Alvin, M. A., Lippert, T. E., Diaz, E. D., and Smeltzer, E. E., 1996, ''Filter Component Assessment,'' *Proceedings of the Advanced Coal-Fired Power Systems' 96 Review Meeting*, Morgantown, WV, DOE/METC-96/037.

[4] Swanson, M. L., and Ness, R. O., 1996, ''Hot-Gas Filter Testing with a Transport Reactor Development Unit,'' *Proceedings of the Advanced Coal-Fired Power Systems' 96 Review Meeting*, Morgantown, WV, DOE/METC-96/1037.

[5] Swanson, M. L., 1998, personal communications, July.

Reduced Mechanisms for Prediction of NO_2 Formation and Ignition Delay in Methane-Air Combustion

R. Homma
Energy and Environmental Technology Laboratory,
Tokyo Gas Co. Ltd.,
16-25 Shibaura, 1-Chome Minato-ku,
Tokyo, 105-0023 Japan

J.-Y. Chen
Department of Mechanical Engineering,
University of California at Berkeley,
Berkeley, CA 94720-1740

Two new 14-step and 16-step reduced mechanisms for methane-air combustion were systematically developed by assuming the quasisteady state for 26–28 species in the starting mechanism. A series of comparison between the reduced mechanisms and the starting mechanism was carried out with the emphasis on their capabilities in predicting NO_2 formation and ignition delay. The two reduced mechanisms successfully capture the complex behaviors of NO_2 formation, which depends on the characteristic mixing time, pressure, and the contamination of hydrocarbon in air. The flame structure and NO_x formation in diffusion flame were well predicted by the 16-step mechanism, while the 14-step showed less satisfactory performance on predicting prompt NO formation. The 16-step mechanism was shown accurate in predicting ignition delay over a wide range of equivalence ratio, temperature and pressure. The necessity of including CH_2O, C_2H_6, C_2H_4, and HO_2 in the reduced mechanisms was discussed. [DOI: 10.1115/1.1360687]

Introduction

Nitrogen dioxide (NO_2) emission from combustion devices has been an important issue because of its direct health effect and its impact on the formation of photochemical oxidants. A considerable amount of NO_2 has been observed in the Gas Turbine (GT) exhaust ([1]). One of the NO_2 formation mechanisms is through mixing of hot gases containing NO with the cooling air in combustors ([2,3]). A significant amount of NO_2 can also form in the cooling process from the combustor to the stack exit ([4]). Therefore it is essential to properly design cooling and dilution processes in and after the combustor to minimize NO_2 emission from GT's. Autoignition of the fuel-air mixture during premixing has been a concerned issue in lean premixed combustor in high performance GT's ([5]). Since autoignition and NO_2 formation are influenced strongly by the interactions between the flow field and the chemical reactions, an accurate model for chemical reactions is essential for using CFD as a design tool. Unfortunately, direct implementation of a large reaction mechanism into a three-dimensional CFD is computationally prohibited at the present time.

Recently, several approaches have been proposed for reducing computational tasks of chemical reactions. The in situ adaptive tabulation (ISAT) creates an adaptive data structure that stores the results obtained from direct integration of stiff chemical kinetics. Once the data is established, ISAT allows retrieval of changes in chemical states with a prescribed accuracy ([6]). Since the CPU time used in retrieving the stored data is much smaller than that needed to perform a direct integration, ISAT significantly reduces the amount of computational time for chemical kinetics. Another approach is to utilize artificial neural networks (ANN) for a nonlinear interpolation of the results from direct integration of chemical kinetics ([7,8]). Both ISAT and ANN were found to be most effective when the reaction mechanism contains a small number of species. Therefore it is desirable to develop reduced chemical mechanisms that are accurate in predicting many aspects of combustion phenomena with a minimal number of species.

Two new reduced mechanisms with 18 and 20 species have been systematically developed assuming the quasisteady state for 26–28 species in the starting mechanism. The assessment of the reduced mechanisms will be focused on their performance in predicting NO_2 formation and ignition delay. The present comparison will be limited to their capability in reproducing the characteristics of the starting mechanism since the reliability of the starting detailed reaction is not addressed here.

Reduced Mechanisms

The detailed mechanism developed by Miller and Bowman [9] has been used in several studies of autoignition and NO_2 formation in GT's ([10,11,4]). This detailed mechanism contains 51 species and 240 steps. Exploration runs showed that C_3 and C_4 chemistry and C_2N_2 are not important and a staring mechanism containing 46 species and 224 reactions is used.

The development of the reduced mechanisms was performed through the following three steps. First, several perfectly stirred reactor (PSR) calculations and the corresponding sensitivity analysis were performed to generate information on the species to be assumed in the quasisteady state. The inlet gas to the PSR was a self-ignitable mixture with a small amount of heat release (CH_4: 0.1 percent, C_2H_6: 0.1 percent, CO: 0.86 percent, H_2: 1.72 percent, O_2: 3.0 percent, CO_2: 7.72 percent, H_2O: 15.44 percent, N_2: 71.0 percent, $T_{in}=800$ K, $P=1$ atm). Five runs with different residence times were conducted: $\tau=10$ ms, $T=837$ K; $\tau=44$ ms, $T=917$ K; $\tau=100$ ms, $T=1038$ K; $\tau=1$ sec, $T=1050$ K; $\tau=10$ sec, $T=1052$ K.

Second, the relative and the absolute QSS errors defined by

$$\delta_{\text{relative}}=\frac{|\dot{\omega}_k^p-\dot{\omega}_k^c|}{\max(|\dot{\omega}_k^p|,|\dot{\omega}_k^c)},$$

$$\delta_{\text{absolute}}=X_k\frac{|\dot{\omega}_k^p-\dot{\omega}_k^c|}{\max(|\dot{\omega}_k^p|,|\dot{\omega}_k^c|)}$$

were computed, where $\dot{\omega}_k^p$ and $\dot{\omega}_k^c$ are the production rate and the consumption rate of kth species and X_k is the mole fraction of kth species. Using this information, a group of species in the starting

Contributed by the Combustion and Fuels Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the C&E Division, October 2000; final revision received by the ASME Headquarters December 2000. Editor: H. D. Nelson.

Table 1 The 16-step reduced mechanism and the 14-step mechanism excluding (R13) and (R14).

C_1 reactions	(R1)	$CH_4+OH=CH_3+H_2O$
	(R2)	$CH_3+O=CH_2O+H$
	(R3)	$CH_2O+O=CO+H_2O$
	(R4)	$CO+OH=CO_2+H$
C_2 reactions	(R5)	$2CH_3=C_2H_6$
	(R6)	$C_2H_6=C_2H_4+H_2$
	(R7)	$C_2H_4=C_2H_2+H_2$
H-O reactions	(R8)	$H+O_2=O+OH$
	(R9)	$O+H_2=H+OH$
	(R10)	$H+H_2O=OH+H_2$
	(R11)	$H+OH=H_2O$
HO_2-H_2O_2 reactions	(R12)	$H+O_2=HO_2$
	(R13)	$2HO_2=H_2O_2+O_2$
NO_x reactions	(R14)	$N_2+CH_3+O_2=HCN+NO+H_2O$
	(R15)	$N_2+O=NO+N$
	(R16)	$NO+HO_2=NO_2+OH$

mechanism can be identified in good approximation by the QSS. Once species in QSS were selected, a reduced mechanism is constructed by an automatic matrix operation ([12,13]). Third, the candidate mechanism was tested mainly by the transient well-mixed reactor (TWMR) as described in the following section ([14]). These procedures were repeated until the appropriate reduced mechanism was developed.

Through the above procedures, 28 species (C, C_2H, CH, CH_2(s), NNH, CN, N, H_2CN, HCCO, NH, CH_2, NH_2, HCO, NCO, C_2H_3, CH_2OH, NH_3, CH_3O, C_2H_5, HOCN, CH_2CO, HNO, HCCOH, HNCO, HCNO, N_2O, H_2O_2, and HCN) were identified in QSS leading to a 16-step reduced mechanism. In addition, a 14-step reduced mechanism was developed by applying QSS to H_2O_2 and HCN for exploring their influence. These two reduced mechanisms are summarized in Table 1.

Assessment of the Reduced Mechanisms

NO_2 Formation. The TWMR is an idealized continuous flow reactor which allows the growth of the internal mass due to continuous inflow streams ([14]). Since the mixing in the reactor is assumed to be infinitely fast, the mixture in the TWMR is treated as molecularly uniform. The dilution of combustion products by cold air was simulated by the TWMR as follows. At the initial state, the TWMR contains high-temperature combustion products of stoichiometric methane-air mixture ($T=2225$ K, $P=1$ atm, equilibrium composition for those species included in the 14-step mechanism except NO and NO_2, NO=300 ppm, $NO_2=0$ ppm). A continuous injection of air at ambient temperature was conducted to mimic the dilution process. The inflow rate of air was set such that the mass of the TWMR doubles during each mixing characteristic time τ_{mix}. This mixing procedure is closely related to the binary mixing models used in Monte Carlo simulations of turbulent combustion where two randomly sampled particles of same mass mix and produce two particles in intermediate composition ([15]). In order to cover a wide range of mixing time scales, τ_{mix} was varied from 0.1 to 100 ms. A typical temperature history during the TWMR calculation is presented in Fig. 1.

Figure 2 shows the evolution of NO_2/NO_x ratio ($NO_x=NO+NO_2$) as a function of mixture temperature during the dilution process at ambient pressure. As the mixture temperature reduces to about 1500 K, conversion of NO into NO_2 initiates. When the combustion products gradually mix with air, as in the case of $\tau_{mix}=100$ ms, the amount of NO_2 formed is small. However, if the dilution is fast as in the case of $\tau_{mix}=0.1$ ms, the amount of conversion becomes significant. One interesting feature is that the temperature at which NO_2 starts to form shifts to a lower value as the mixing becomes faster. As seen in Fig. 2, both the 14-step and the 16-step mechanisms produced nearly identical results to those obtained with the starting mechanism for $\tau_{mix}=0.1$ to 100 ms. Figure 3 presents similar results of NO_2/NO_x ratio during the dilution process at 30 atm. The final NO_2/NO_x ratios exhibit a nonlinear dependence on the mixing characteristic time that is reproduced successfully by the two reduced mechanisms.

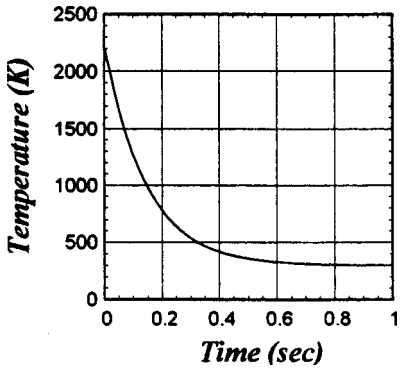


Fig. 1 Typical temperature history of the TWMR calculation ($\tau_{mix}=100$ ms)

It has been known for some time that addition of a small amount of hydrocarbon to the post-flame gas substantially promotes the NO-NO_2 conversion ([16,17]). This phenomenon is important for understanding and mitigating NO_2 emissions from practical devices. Figure 4 presents the predicted NO_2/NO_x ratios during a dilution process with air doped by 500 ppm of CH_4. Comparisons of Figs. 2 and 4 reveal that addition of CH_4 promotes NO_2 conversion most noticeable for slow mixing cases.

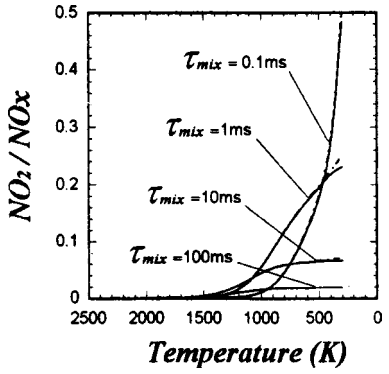


Fig. 2 Evolution of NO_2/NO_x ratio in the TWMR during the dilution by air at ambient pressure. Lines: Miller-Bowman [9], broken lines: 14 step; dotted lines: 16 step.

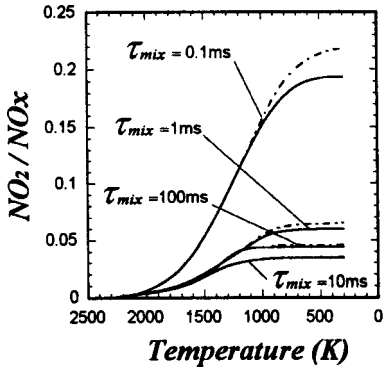


Fig. 3 Evolution of NO_2/NO_x ratio in the TWMR during the dilution by air at elevated pressure ($p=30$ atm). Lines: Miller-Bowman [9], broken lines: 14 step; dotted lines: 16 step.

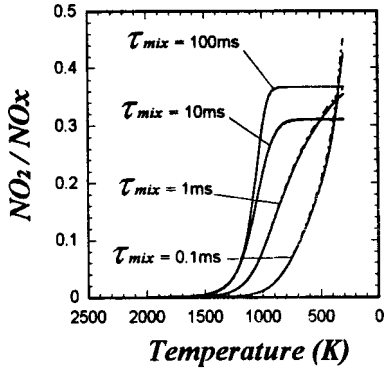


Fig. 4 Evolution of NO_2/NO_x ratio in the TWMR during the dilution by air doped with 500 ppm of methane at ambient pressure. Lines: Miller-Bowman [9], broken lines: 14 step; dotted lines: 16 step.

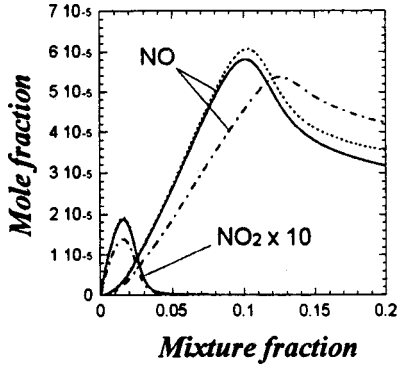


Fig. 7 Profiles of NO and NO_2 mole fractions in opposed methane-air diffusion flame with strain rate $a=100$/s. Solid lines: Miller-Bowman [9], broken lines: 14 step; dotted lines: 16 step.

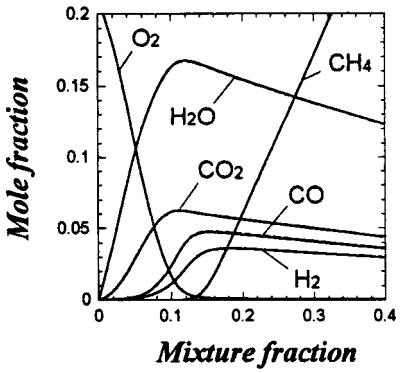


Fig. 5 Major species profiles in methane-air opposed flame. Strain rate $a=100$ (1/s). Lines: Miller-Bowman [9], broken lines: 14 step; dotted lines: 16 step.

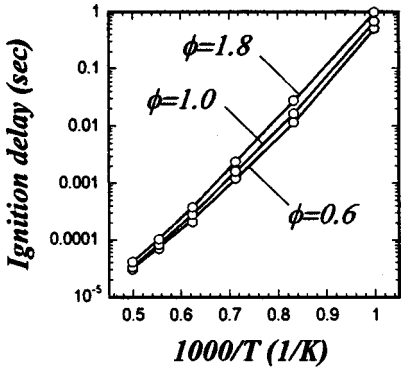


Fig. 8 Dependence of ignition delay time on initial mixture temperature at $p=1$ atm. Lines: Miller-Bowman [9], symbols: 16-step reduced mechanism.

This feature complicates the dependence of NO_2 emission on the characteristic mixing time in the presence of a small amount of hydrocarbons. The two reduced mechanisms are found accurate in predicting this feature.

Methane-Air Opposed Diffusion Flame. Using opposed diffusion flames, we further examine different aspects of the validity of the reduced mechanisms because: (1) molecular diffusion has an important role, and (2) reaction is accompanied with a significant heat release. The opposed laminar diffusion flame of methane and air was computed for conditions at $P=1$ atm, $T_{CH_4}=300$ K, $T_{air}=300$ K, and at a strain rate, $a=100$/s. Figure 5 shows the comparison of the flame structure obtained with the two reduced mechanisms and the starting mechanism. Both the 14-step and 16-step mechanisms successfully reproduced the major species profiles obtained with the starting mechanism. The two reduced mechanisms were also accurate in predicting O, OH, H, and HO_2 concentrations as shown in Fig. 6. Figure 7 shows the comparison of NO and NO_2 concentrations. The 16-step mechanism showed good agreement with the starting mechanism; however, the 14-step mechanism was less accurate in predicting both NO and NO_2 concentrations.

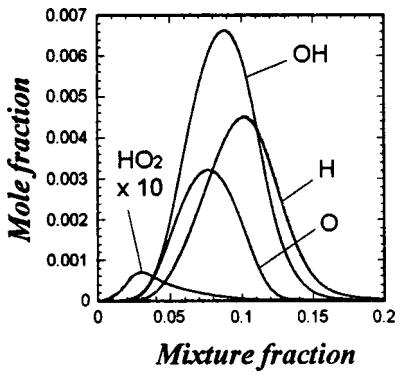


Fig. 6 H, O, OH, and HO_2 profiles in methane-air opposed flame. Strain rate $a=100$/s. Lines: Miller-Bowman [9], broken lines: 14 step; dotted lines: 16 step.

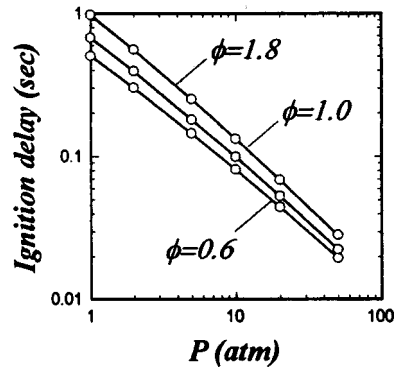


Fig. 9 Dependence on ignition delay on pressure with initial mixture at $T=1000$ K. Lines: Miller-Bowman [9], symbols: 16-step reduced mechanism.

Ignition Delay. Ignition delay time was examined using a plug flow reactor model with methane-air mixtures of different equivalence ratios (ϕ). Figure 8 shows the dependence of the ignition delay time on the initial mixture temperature at atmospheric pressure obtained with the starting mechanism and the 16-step reduced mechanism. Figure 9 presents the dependence of ignition delay time on pressure at a fixed initial temperature of 1000 K. As seen in Figs. 8 and 9, the 16-step reduced mechanism is accurate in predicting ignition delay for a wide range of equivalence ratio, temperature, and pressure.

Discussion

Effect of QSS Assumption on C_2 Species. The two reduced mechanisms developed here contain CH_2O, HO_2, and several C_2 species, many of which were not included in the previous reduced mechanisms ([18,19,20]). To understand the relative importance of these species, a comparative study was performed using several 15-step reduced mechanisms. Figure 10 shows the temperature histories during autoignition of a lean methane-air mixture (at $\phi=0.6$) obtained by the starting mechanism, the 16-step mechanism, and several 15-step mechanisms. These 15-step mechanisms were based on the 16-step reduced mechanism but assuming QSS for one additional species (15 step-a assumed QSS for CH_2O; 15 step-b assumed QSS for C_2H_6; 15 step-c assumed QSS for C_2H_4; 15 step-d assumed QSS for HO_2). As shown in Fig. 10, the QSS assumption for CH_2O, C_2H_6, or C_2H_4 resulted in shorter ignition delay times than those obtained with the starting mechanism and the 16-step mechanism. On the other hand, if we assume QSS for HO_2, the ignition delay was overestimated. A similar trend was also seen for rich mixtures (not shown). It is noteworthy that the QSS assumption for either C_2H_4 or C_2H_6 species deteriorates the prediction of ignition delay even in very lean conditions.

The detailed roles of CH_2O, C_2H_6, C_2H_4, and HO_2 during pre-ignition was analyzed with the aid of a carbon atom flow chart. Figure 11 shows the relative magnitudes of carbon atom fluxes during pre-ignition of a lean methane-air mixture ($\phi=0.6$) at $T=1100$ K. In the later stage of methane decomposition during pre-ignition, HO_2 was formed through the reaction step (A) $HCO+O_2=HO_2+CO$ leading to the promotion of methane oxidation through (B) $CH_3+HO_2=CH_3O+OH$ and (C) $CH_4+OH=CH_3+H_2O$. As the carbon atom flux of reaction step (A) increases, a shorter ignition delay is predicted.

As illustrated in Fig. 11, the production of CH_2O, C_2H_6, C_2H_4, and CO is larger than their consumption during pre-ignition and hence these species accumulate. If we assume QSS for these species, the accumulation of these species is set zero leading to overestimation of the flux of reaction (A); consequently, a shorter ignition delay is predicted. If HO_2 is assumed to be in QSS, the concentration of HO_2 is underestimated and a longer ignition delay is predicted. In short, it is necessary to include CH_2O, C_2H_6, C_2H_4, and HO_2 in the reduced mechanism for accurate predictions of ignition delay for both rich and lean mixtures.

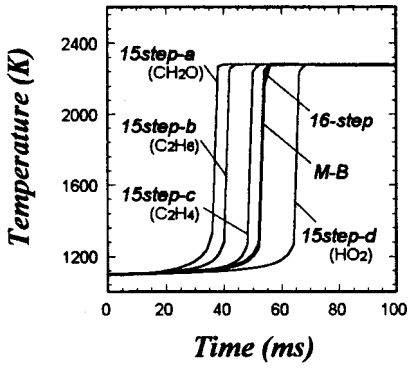


Fig. 10 Temperature profiles during autoignition process obtained with various reduced mechanisms (thin lines) and the starting detailed mechanism (thick line) at lean condition ($\phi=0.6$). 15 step-a: assumes QSSA for CH_2O; 15 step-b: assumes QSSA for C_2H_6; 15 step-c: assumes QSSA for C_2H_4; 15 step-d: assumes QSSA for HO_2. "M-B" denotes Miller-Bowman [9] mechanism.

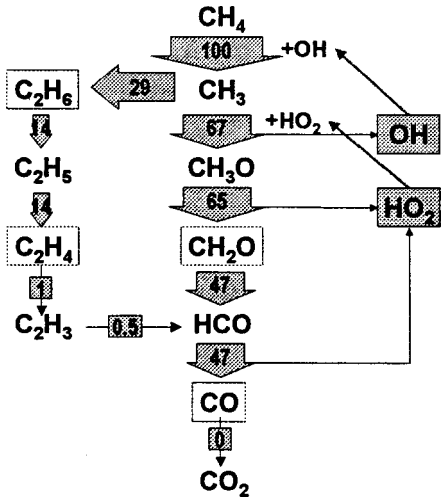


Fig. 11 Carbon atom flow diagram during pre-ignition process of methane-air mixture ($\phi=0.6$, $T_{in}=1100$ K)

NO_2 Formation Mechanisms. As shown in Fig. 2, fast mixing of combustion products with air promotes NO to NO_2 conversion. However, when a small amount of hydrocarbon is contained in the air, NO_2 formation is substantially enhanced in slow mixing as presented in Fig. 4. These opposite dependencies of NO_2 formation on the characteristic mixing time stem from two different NO_2 formation mechanisms.

The dominant reaction mechanism responsible for NO_2 formation is via (1) $NO+HO_2=NO_2+OH$. This reaction has a negative dependence on temperature. This combines with the high level of HO_2 at low temperature leads to rapid NO_2 formation. There are two important paths to form HO_2 during dilution/cooling process: (2) $H+O_2+M=HO_2+M$ and (3) $HCO+O_2=HO_2+CO$.

In case of mixing with air without hydrocarbon, the enhanced H radical pool by fast cooling is the main cause for rapid NO_2 formation. Figure 12 shows the ratios of H, O, and OH in the fast mixing case with $\tau_{mix}=1$ ms to those in the slow mixing case with $\tau_{mix}=100$ ms. As seen in Fig. 12, fast cooling leads to high levels of H, O, and OH radicals at low temperatures because the recombination process of these radicals is slow at low temperatures. In

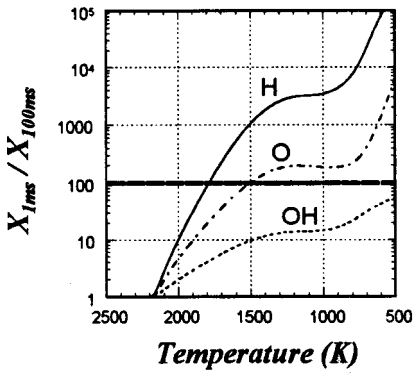


Fig. 12 Ratios of H, O, and OH concentration in fast mixing case ($\tau_{mix}=1$ ms) to those in slow mixing case ($\tau_{mix}=100$ ms) showing a significant buildup of H radical when mixing is fast

addition, the H radical is increased significantly through the reaction sequences: (4) $O+H_2O=2OH$ and (5) $CO+OH=CO_2+H$. As a result, fast cooling promotes NO_2 formation through reaction (2) and (1). When air is contaminated with hydrocarbons, reaction (3) becomes the dominant step in HO_2 formation. In this case, the level of HO_2 is determined by HCO concentration, which depends on the amount of hydrocarbons and the residence time. Therefore a large amount of NO was converted to NO_2 in slow mixing cases. Both of these two different paths of NO_2 formation were properly included in the two reduced mechanisms.

Conclusions

Two new reduced mechanisms for methane-air combustion were developed. The validity of the reduced mechanisms were extensively examined for NO_2 formation during dilution process of combustion products, opposed diffusion flames, and ignition delay time. The complicated dependence of NO_2 formation on the mixing characteristic time and hydrocarbon contaminated in air were successfully reproduced by both 14-step and 16-step mechanisms. The flame structure and NO_x formation in diffusion flames were well predicted by 16-step mechanism. The 14-step mechanism was found less accurate in predicting NO formation. The 16-step mechanism was shown accurate for predicting ignition delay over a wide range of equivalence ratio, temperature, and pressure. The analysis of carbon atom flow diagram during preignition revealed the importance of CH_2O, C_2H_6, C_2H_4, and HO_2 and they should be included in the reduced mechanism.

Acknowledgments

The first author is indebted to University of California at Berkeley for providing a visitor post. The first author would also like to thank Professor Dibble, UC Berkeley, for co-hosting.

References

[1] Johnson, G. M., and Smith, M. Y., 1978, ''Emissions of Nitrogen Dioxide from a Large Gas-Turbine Power Station,'' Combust. Sci. Technol., **19**, 67–70.
[2] Sano, T., 1984, ''NO_2 Formation in the Mixing Region of Hot Burned Gas with Cool Air,'' Combust. Sci. Technol., **38**, pp. 129–144.
[3] Hori, M., 1986, ''Experimental Study of Nitrogen Dioxide Formation in Combustion Systems,'' Twenty-first Symposium (International) on Combustion, pp. 1181–1188.
[4] Feitelberg, A. S., and Correa, S. M., 1999, ''The Role of Carbon Monoxide in NO_2 Plume Formation,'' ASME paper No. 99-GT-053, pp. 1–7.
[5] Correa, S. M., Dean, A. J., and Hu, I. Z., 1996, ''Combustion Technology for Low-Emission Gas-Turbines: Beyond NO_x,'' ASME J. Energy Resour. Technol., **118**, pp. 193–200.
[6] Pope, S. B., 1997, ''Computationally Efficient Implementation of Combustion Chemistry using In Situ Adaptive Tabulation,'' Combust. Theory Modell., **1**, pp. 41–63.
[7] Blasco, J. A., Fueyo, N., Larroya, J. C., Dopazo, C., and Chen, J.-Y., 1999, ''A Single-Step Time-Integrator of a Methane-Air Chemical System using Artificial Neural Networks,'' Comput. Chem. Eng., **23**, pp. 1127–1133.
[8] Blasco, J. A., Fueyo, N., Dopazo, C., and Chen, J.-Y., 2000, ''A Self-Organized-Map Approach to Chemistry Representation in Combustion Applications,'' Combust. Theory Modell., **4**, pp. 61–76.
[9] Miller, J. A., and Bowman, C. T., 1989, ''Mechanism and Modeling of Nitrogen Chemistry in Combustion,'' Prog. Energy Combust. Sci., **15**, pp. 287–338.
[10] Correa, S. M., 1994, ''Models for High-Intensity Turbulent Combustion,'' Comput. Syst. Eng., **5**, No. 2, pp. 135–145.
[11] Amano, T., and Hase, K., 1994, ''Cooling Conditions of Hot Exhaust Gas for Low Conversion of NO to NO_2,'' J. Inst. Energy, **67**, pp. 174–180.
[12] Chen, J.-Y., 1988, ''A General Procedure for Constructing Reduced Reaction Mechanisms with Given Independent Relations,'' Combust. Sci. Technol., **57**, pp. 89–94.
[13] Chen, J.-Y., 1997, ''Development of Reduced Mechanisms for Numerical Modeling of Turbulent Combustion,'' Workshop on ''Numerical Aspects of Reduction in Chemical Kinetics,'' CERMICS-ENPC Cite Descartes—Champus sur Marne, France, pp. 1–23.
[14] Wang, Z., and Chen, J.-Y., 1997, ''Modeling of Microscale Turbulence and Chemistry Interaction in Near-Field Aircraft Plumes,'' J. Geophys. Res., **102**, pp. 12,871–12,883.
[15] Pope, S. B., 1985, ''PDF Methods for Turbulent Reactive Flows,'' Prog. Energy Combust. Sci., **11**, pp. 119–192.
[16] Sano, T., 1985, ''NO_2 Formation in the Mixing Region of Hot Burned Gas with Cool Air—Effect of Surrounding Air,'' Combust. Sci. Technol., **43**, pp. 259–269.
[17] Hori, M., Matsunaga, N., Malte, P. C., and Marinov, N. M., 1992, ''The Effect of Low-Concentration Fuels on the Conversion of Nitric Oxide to Nitrogen Dioxide,'' 24th Symp. (International) on Combustion, pp. 909–916.
[18] Peters, N., and Kee, R. J., 1987, ''The Computation of Stretched Laminar Methane-Air Diffusion Flames Using a Reduced Four-Step Mechanism,'' Combust. Flame, **68**, pp. 17–29.
[19] Smooke, M. D., 1991, ''Reduced Kinetic Mechanisms and Asymptotic Approximations for Methane-Air Flames,'' Lecture Notes in Physics, 384, Springer-Verlag, pp. 1–28.
[20] Peters, N., and Rogg, B., ed., 1993, ''Reduced Reaction Mechanisms for Applications in Combustion System,'' Lecture Notes in Physics, m15, Springer-Verlag.

An Experimental Study of Surface Temperature Distribution on Effusion-Cooled Plates

K. M. Bernhard Gustafsson

T. Gunnar Johansson
e-mail: gujo@tfd.chalmers.se

Department of Thermo and Fluid Dynamics,
Chalmers University of Technology,
SE-412 96, Göteborg, Sweden

A parametric study of temperature distribution on effusion-cooled plates under conditions typical for combustion chambers was performed using infrared thermography. In this investigation, the effects of different temperature ratios, velocity ratios of the two air streams, the injection hole spacing, inclination angle of the injection holes, and the thermal heat conductivity of the plates were studied. For a given amount of cooling air, the cooling efficiency was found to increase markedly with a reduction in hole spacing, i.e., when the number of holes was increased. Reducing the injection angle results in more attached jets, especially for small amounts of cooling air, and marginally lowers the wall temperature. A high thermal conductivity of the plate was found to decrease its surface temperature in front of the first row of holes but not the mean temperature in downstream positions. The most important operational parameters were the temperature ratio and the velocity ratio of the hot and cold air streams. An almost linear relation was found between the temperature ratio and the surface temperature when the jet velocity was large compared to the crossflow velocity. For plates with sparse hole spacing, a change in the velocity ratio had a small effect on the surface temperature, whereas the effect was large for dense hole spacings and the same amount of cooling air.
[DOI: 10.1115/1.1364496]

1 Introduction

In modern gas turbines, different parts of the combustion chamber, the turbine, and flame holder become hot and need efficient cooling. Without proper cooling, the lifetime of these parts is drastically shortened. The temperature of the exposed areas may even exceed the melting point of the metal if not cooled. Increasing the efficiency of gas turbines is coupled with higher temperature and pressure in the combustion chamber, which further calls for high performance cooling methods. Effusion cooling (full coverage film cooling) is one powerful way to protect the hot walls.

The short-term goal of this project was to provide experimental temperature data useful for the design of effusion-cooled combustion chamber walls. A parametric study of the temperature distribution of effusion-cooled plates was therefore performed. The experiments were scaled to match real combustion chamber conditions. It is intended that the results of this study will be helpful in finding designs in which it is ascertained that the wall temperature is kept below a certain value for the varying operational parameters at hand, to be accomplished with a minimum amount of cooling air. Many different operational and design parameters exert an influence on the surface temperature distribution. This study was intended to examine the way in which these parameters act and their relative importance. All temperature data from these measurements, comprising 242 different cases with different design and operational parameters, are given in the report by Gustafsson [1]. These results might also be useful for comparison with numerical results.

A large number of investigations on film cooling have been carried out and the research area is very active. This study differs from most others in that here the cooling effect of several rows of holes, distributed over 100 diameters, is studied. The results of investigations on the effect of single or double rows of holes cannot be adopted to this case with several rows of holes. Vortices generated by different jets in different rows will interact with each other, creating flow patterns that are not present in the single row cases.

Measurements of several rows-of-holes film cooling with infrared thermography were reported by, e.g., Martiny et al. [2]. Heat transfer measurements using a mass transfer analogy have been reported by several authors, e.g., Cho and Goldstein [3,4]. Foster and Lampard [5], and Friedrichs et al. [6]. Temperature measurements on walls with low heat conductivity were reported by, e.g., Gritsch et al. [7], van Treuren et al. [8], and Salcudean et al. [9]. Adiabatic and isoenergetic Stanton number measurements have been reported by, e.g., Ligrani and Ramsey [10]. Methods to predict heat transfer in film-cooled systems from adiabatic and isoenergetic heat transfer coefficients are given in Eckert [11]. A review of methods for calculations of heat transfer under film cooling is given in Leontiev [12].

2 Experiment

The wall temperatures were measured with an infrared camera. The camera records the sum of heat radiation from graybody emission from the test plate and reflections from the surroundings. As side walls were hotter than the cooled test plate, the measured wall temperature was corrected by subtracting the diffuse reflections via a simple radiative heat transfer model. An error estimate of the measured wall temperatures is presented below.

In the experiments, the temperature ratio, T_0/T_c, between the hot and cold air streams was varied by keeping the cold air at 30–40°C and varying the hot air temperature. Three different hot air temperatures were chosen, 200, 250, and 300°C. When large amounts of cooling air were supplied in the 300°C cases, the heater was unable to deliver the power needed. Under these circumstances a hot air temperature of 150°C was used. Three levels of hot air stream velocity were used: 17, 30, or 46 m/s. The cold air stream velocity levels in the plenum chamber (see Fig. 2) were 0.3, 0.5, or 0.7 m/s.

2.1 The Wind Tunnel. The wind tunnel consists of one hot main loop and one cold secondary closed loop; see Fig. 1. The main loop is able to withstand temperatures of 300°C and is

Contributed by the Combustion and Fuels Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the C&F Division, May 2000; final revision received by the ASME Headquarters Jan. 23, 2001. Editor: H. D. Nelson.

heated by an 8-kW electrical heater. The turbulence level was 1 percent in the free-stream according to LDA measurements. Air for the secondary loop is taken from the main loop downstream of the test section and is cooled to 30°C by a compact heat exchanger. The wind tunnel is operated at normal air pressure.

The size of the test section is 80×80×720 mm. It allows exchange of the test plates and the inner side walls. Several thermocouples are mounted in the surrounding walls as the surrounding temperature is needed for the surface temperature correction procedure; see Fig. 2. Optical access is made possible via six openings in the test section lid.

2.2 The Test Plates. The size of the test plates was 80×720 mm, with a thickness b of 20 mm. In all experiments, the holes had a staggered placement and a diameter d of 5 mm (see Fig. 3), allowing four lateral holes. As the effect of different hole spacings $\delta x/d$, injection angles α, and thermal heat conductivity ratios λ_s/λ_c was studied, a total of ten test plates was manufactured, nine in teflon ($\lambda_s = 0.5\,\mathrm{Wm^{-1}K^{-1}}$) and one in steel ($\lambda_s = 45\,\mathrm{Wm^{-1}K^{-1}}$). The teflon plates had three different hole spacings $\delta x/d$, 3, 6, or 12, and three different injection angles α, 15, 20, or 30 deg. For the steel plate case $\delta x/d = 6$ and $\alpha = 30$ deg.

2.3 Wall Temperature Measurements. Infrared thermography was chosen because of its high thermal and spatial resolution. An AGEMA THV 900 LW system working in the 8–12-μm wavelength band was used. The infrared camera used had a 272 ×136 pixel digital output, which could be stored and processed on a workstation. The spatial resolution was 1.5 mrad at 50 percent modulation, which corresponds to 0.75 mm at the focusing distance used. One pixel in the thermographs corresponds to approximately 0.64 mm.

As the whole test plate could be covered with six thermographs, a complete mapping of the test plate was made. The six corrected thermographs overlapped each other by approximately 1 cm. An algorithm for stitching them together was constructed, resulting in seamless joints. The algorithm worked by finding the overlap between neighboring images that minimized the temperature difference and temperature gradient difference in the overlapping areas.

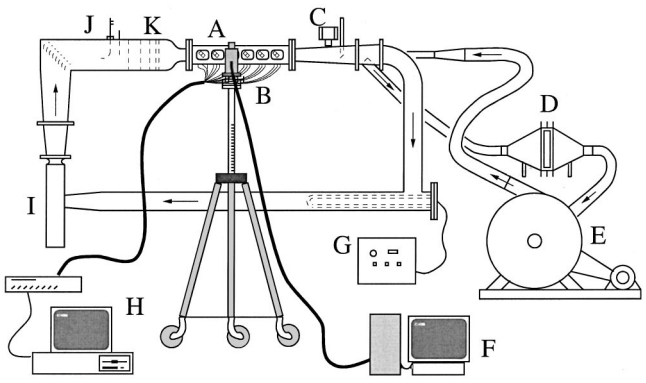


Fig. 1 The wind tunnel. (A) Test section, (B) IR camera, (C) vortex flow-meter, (D) compact heat exchanger, (E) cold air loop fan, (F) IR camera system controller, (G) heater and controller, (H) thermocouple measurement instrument, (I) hot air loop fan, (J) temperature sensor [Pt-100], and (K) screens

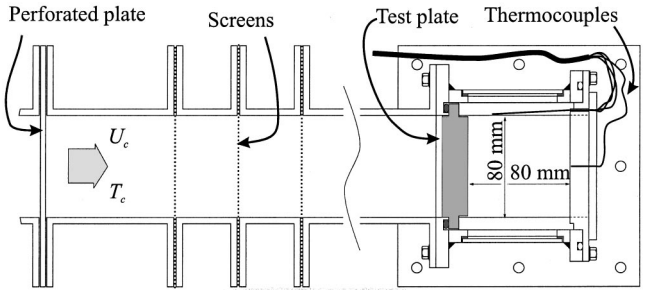


Fig. 2 Cross section of the plenum chamber and wind tunnel

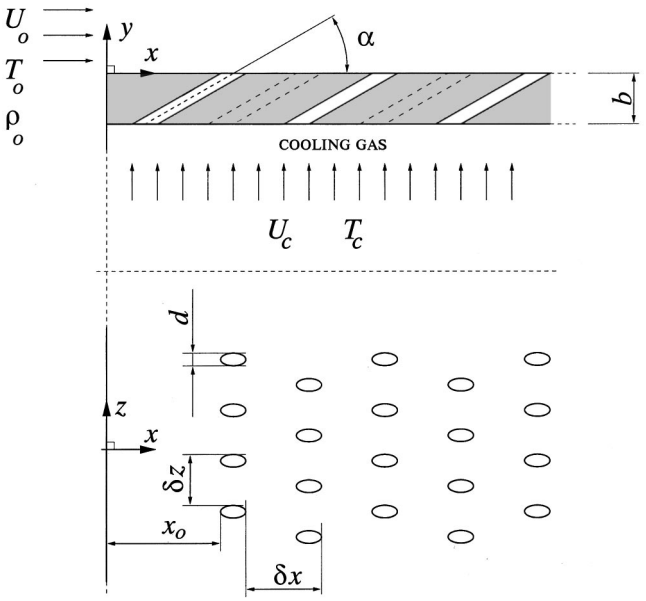


Fig. 3 An idealized case used for the dimensional analysis

2.4 Wall Temperature Correction. The test plates were painted with a NEXTEL velvet coating with high emissivity, 0.96, in order to reduce the reflections from the hot surrounding walls. A correction of the measured temperature in one pixel was made to compensate for the diffuse reflections. The corrected temperature T_1 was calculated as

$$T_1 \approx \left(\frac{T_{\mathrm{IR}}^4}{\varepsilon} - \frac{\rho^D}{A_1} \sum_{i=1}^{m} F_{i1} A_i T_i^4 \right)^{1/4} \tag{1}$$

where T_{IR} is the temperature measured by the camera with an emissivity setting of 1.00. Further, ε is the emissivity of the surface (0.96), ρ^D is the diffuse part of the total reflectivity, A_1 is the measurement area, F_{i1} is the view factor from an area on the surrounding walls, A_i, to A_1, and T_i is the temperature measured by the thermocouples at A_i. In the computations, analytical formulas for the view factor were used for perpendicular surface elements and numerical integration was used for parallel surface elements. The specular reflections are not considered because these are influential only close to the side walls, where the flow is perturbed in any event by boundary layers and secondary motions.

2.5 Wall Temperature Error Estimate. An error estimate was derived from a linearization of Eq. (1). Let ΔT_1 denote the error estimate of T_1, ΔT_{IR} the error in the camera reading, ΔT_i the error in the thermocouple reading, $\Delta\varepsilon$ the error in emissivity, and $\Delta\rho^D$ the error in reflectivity. The estimate given in Eq. (2) is valid if $\Delta T_1/T_1$, $\Delta T_{\mathrm{IR}}/T_{\mathrm{IR}}$, and $\Delta T_i/T_i$ are small compared with unity.

$$|\Delta T_1| \approx \left| \left(\frac{T_{\mathrm{IR}}}{T_1} \right)^3 \frac{\Delta T_{\mathrm{IR}}}{(\varepsilon + \Delta\varepsilon)} \right| + \left| \frac{\rho^D \varepsilon}{T_1^3 (\varepsilon + \Delta\varepsilon) A_1} \sum_{i=1}^{m} F_{i1} A_i T_i^3 \Delta T_i \right| + \left| \frac{\rho^D \Delta\varepsilon + \varepsilon \Delta\rho^D}{4 T_1^3 (\varepsilon + \Delta\varepsilon) A_1} \sum_{i=1}^{m} F_{i1} A_i (T_i^4 + 4 T_i^3 \Delta T_i) \right| \tag{2}$$

In the temperature error estimate calculations, ΔT_{IR} was, according to the manufacturer of the infrared camera, ± 2.5°C, ΔT_i was ± 1.3°C, and $\Delta \rho^D$ and $\Delta \varepsilon$ were estimated to be 0.02.

Errors $\Delta \rho^D$ and $\Delta \varepsilon$ contributed to an error of about 2°C in the total error $|\Delta T_1|$, which is typically 5°C. It should be mentioned that the error in the temperature reading of the infrared camera is probably constant or changes monotonically with temperature. This implies that differences in temperature readings between different plates or different positions are smaller and amount only to about half the total error.

2.6 Emissivity. Measurements of the emissivity were made with the THV 900 system and at other laboratories. An aluminum canister was painted with NEXTEL velvet coating and an insulating layer was wrapped around it, except for a small area where the measurements were made. The canister contained about 80°C water which was well stirred. Pt-100 resistance thermometers were used to measure the ambient room temperature T_{amb}, which was approximately 23°C as well as the temperature of the hot water T_{H_2O}, with an accuracy of $\sim \pm 0.2$°C. As the aluminum wall was very thin ($\sim$0.2 mm), its surface temperature is estimated to be within ± 0.05°C of that of the hot water. A thermograph was taken with the emissivity set to 1.00, which gives T_{IR}. The emissivity of the surface was calculated as

$$\varepsilon = \frac{(T_{IR}^4 - T_{amb}^4)}{(T_{H_2O}^4 - T_{amb}^4)} \tag{3}$$

This method is likely to yield an accuracy in ε of ± 0.02. The main error source in the ε determination is the error in the infrared camera reading (± 1°C is assumed). All measurements estimated the emissivity to be 0.96. The emissivity of old paint used in the experiments and the emissivity of new paint was checked. No change in emissivity caused by degradation or soiling could be observed.

3 Dimensional Analysis

In order to extract only the most relevant dimensionless variables we consider here an idealized case; see Fig. 3. The injection plate (which in the experiments is the test plate) is infinitely long and infinitely wide and the main gas stream extends to infinity in the direction perpendicular to the injection plate. The basis for the analysis is the continuity, momentum and energy equations in their general compressible form and the energy equation for the heat conduction in the injection plate together with the appropriate boundary conditions. Some of the boundary conditions are given in Fig. 3. In addition to these, the fluid velocity must be zero on the solid wall and the temperature and heat flux must be continuous across the solid fluid interface. One thing to be noted here is that we must not specify the inlet density of the hot mainstream gas because the two gas streams must have the same pressure (apart from a small difference due to the pressure loss in the passage of the injection holes). The ideal gas law, assumed to hold here for the gas, thus gives

$$\rho_0 = \rho_c T_c / T_0 \tag{4}$$

We choose here to use as primary scaling variables U_c, T_c, ρ_c, and d. In addition, we assume all material properties of the fluid[1] and wall material to be constant. In a real case, different gases, e.g., combustion gases, with different material properties, as well as the non-negligible radiative heat transfer, may have to be included in the analysis.

The dependent variable in which we are interested here is the surface temperature of the injection plate, T_w. Since we have chosen T_c as a primary scaling variable, we get T_w/T_c, which will be determined as a function of the set of independent variables as found in our dimensional analysis.

From the momentum equation we have a dimensionless viscosity

$$\frac{\mu_c}{\rho_c U_c d} = \frac{1}{\mathrm{Re}_c} \tag{5}$$

where Re_c is the Reynolds number. In the energy equation, a dimensionless thermal conductivity of the fluid appears,

$$\frac{\lambda_c T_c}{\rho_c U_c^3 d} = \frac{1}{(\kappa - 1)\mathrm{Ma}_c \,\mathrm{Re}_c \,\mathrm{Pr}_c} \tag{6}$$

where Pr_c is the Prandtl number, Ma_c is the Mach number, and κ is the specific heat ratio. From the energy equation for the solid we can also extract λ_s/λ_c. In addition to these parameters emerging from the governing equations, we have dimensionless parameters from the boundary conditions: b/d, $\delta x/d$, $\delta z/d$, α, U_0/U_c, and T_0/T_c. Note that the common ways to express blow factors are U_{jet}/U_0, $(\rho U)_{jet/}/(\rho U)_0$, or $(\rho U^2)_{jet}/(\rho U^2)_0$ but, since the two last ones did not arise in the analysis, they were not used. A more extensive derivation is given in Gustafsson [1].

4 Results

Figures 4 and 5 show the surface temperature distribution T_w/T_c for three different velocity ratios U_0/U_c. A unique number is associated to each thermograph, which can be used to cor-

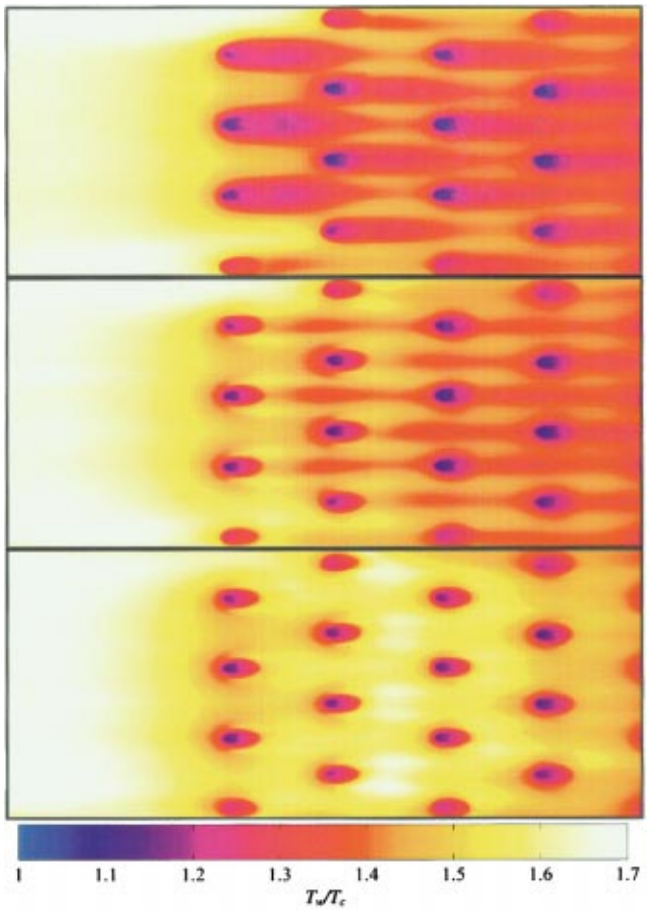


Fig. 4 Closeups of the first rows of holes for teflon plates with $T_0/T_c = 1.78$, $\mathrm{Re}_c = 150$, $\delta x/d = 6$, $\alpha = 30$ deg. A variation in U_0/U_c is made, from the top $U_0/U_c = \{91,60,34\}$. The number of injection holes n is 80.

[1]If the constant in Sutherland's law is small compared with T_c, a proper scaling is achieved even if the viscosity or heat conductivity is temperature dependent.

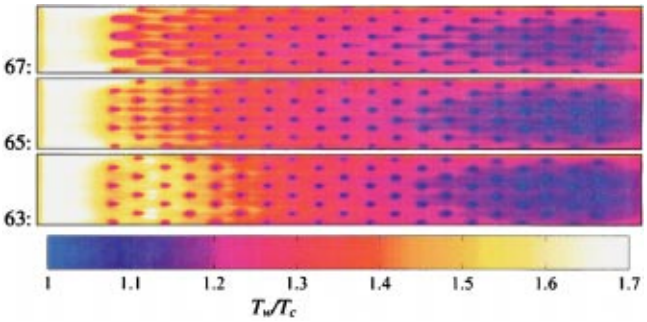


Fig. 5 The effect of different velocity ratios U_0/U_c is shown in thermographs of teflon plates where $T_0/T_c=1.78$, $Re_c=150$, $\delta x/d=6$, $\alpha=30$ deg, and $n=80$. A variation in U_0/U_c is made, from top $U_0/U_c=\{91,60,34\}$.

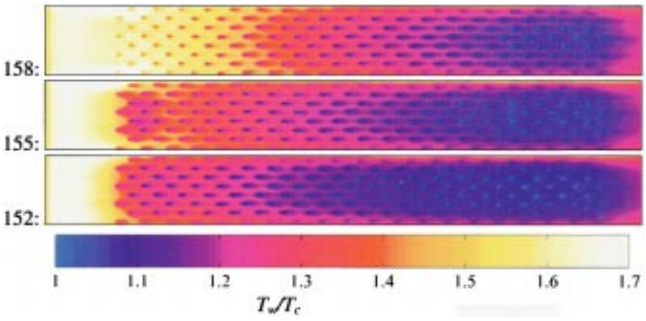


Fig. 6 The effect of different velocity ratios U_0/U_c is shown in thermographs of teflon plates where $T_0/T_c=1.7$, $Re_c=90$, $\delta x/d=3$, $\alpha=20$ deg, and $n=156$. A variation in U_0/U_c is made, from top $U_0/U_c=\{157,102,58\}$.

relate T_w/T_c in Figs. 14–24. Table 1 lists all experimental parameters for the thermographs in this paper. The top thermograph in Fig. 4, in which U_0/U_c is large, shows attached jets, resulting in streaks of low temperature. There is a decrease in temperature in front of the first row of holes resulting from heat conduction in the plate. This effect becomes more pronounced as the velocity ratio is lowered. The test plates were made of teflon with a very low heat conductivity in order to come as close as possible to an adiabatic case and to be able to see the interesting features of the temperature distribution. However, we can see in all cases, from Figs. 4–24, that there is a considerable effect of the conductive heat transfer in front of the holes. The middle thermograph in Fig. 4, with somewhat lower U_0/U_c, also shows streaks of low temperature with two hot areas behind the injection holes. This is an indication that hot air is swept down behind the cold jet by a ''horseshoe'' vortex in the mean velocity field. In the bottom thermograph in Fig. 4, which represents the case with the lowest velocity ratio U_0/U_c, the area behind the injection holes is heated and there are no streaks of low temperature. This is consistent with the idea of fully separated jets, and hot fluid is advected around the jet and reaches the wall with an increased wall temperature as a result. Even streaks of elevated temperature behind the injection holes are visible. The effect is increased some rows downstream, where different vortices may interact. For the lowest velocity ratios, the surface temperature upstream of the first row of holes is lower than in the case with higher velocity ratios. This might be a consequence of less convective heat transfer owing to a larger blockage effect. The importance of the different parameters is discussed below. In Figs. 14–24, the maximum surface temperature T_w^{max}/T_c in the span-wise direction z is given for each x value. The T_w^{max}/T_c parameter is the most important to the lifetime of the cooled parts. Note that a position $x/d=0$ corresponds to the leading edge of the injection hole in the first row.

Table 1 Experimental data for each thermograph

Thermo-graph	T_0 [°C]	T_c [°C]	U_0 [m/s]	U_c [m/s]	δx [mm]	α [°]	n	λ_x
13	249	27.5	28.9	0.30	30	30	84	45
47	200.8	25.6	16.9	0.29	30	30	80	0.5
49	200	28	30.2	0.51	30	30	80	0.5
50	200	26.2	30.3	0.29	30	30	80	0.5
59	250	27	30	0.30	30	30	80	0.5
63	270.1	33.5	17	0.50	30	30	80	0.5
65	270.7	32.8	30	0.50	30	30	80	0.5
67	272.0	32.8	45.6	0.50	30	30	80	0.5
83	250	29	30	0.29	15	30	156	0.5
85	250	30.5	46.1	0.49	15	30	156	0.5
97	200	30	17	0.33	60	20	44	0.5
99	200	30	30	0.33	60	20	44	0.5
101	200	29.5	46.1	0.33	60	20	44	0.5
103	247.4	32.3	17	0.33	60	20	44	0.5
105	247.1	31.6	29.9	0.33	60	20	44	0.5
107	246.6	30.1	46	0.33	60	20	44	0.5
115	150	25.6	46.1	0.33	60	20	44	0.5
116	287.8	32.4	46	0.33	60	20	44	0.5
122	200	28.8	30	0.30	30	20	80	0.5
128	250	28.3	17	0.30	30	20	80	0.5
131	250	30.5	30	0.30	30	20	80	0.5
134	250	28.7	46	0.30	30	20	80	0.5
143	200	25.7	17	0.29	15	20	156	0.5
144	200	27.3	30	0.29	15	20	156	0.5
152	250	27.3	17	0.295	15	20	156	0.5
155	249	28	30	0.295	15	20	156	0.5
157	248	30.3	46	0.49	15	20	156	0.5
158	250	28.5	46	0.293	15	20	156	0.5
171	200.1	32	30.1	0.30	60	15	40	0.5
172	199.7	35.7	46.1	0.48	60	15	40	0.5
194	200	26.2	30.1	0.29	30	15	76	0.5

4.1 Effect of the Temperature Ratio. A higher mainstream temperature T_0/T_c gives a higher surface temperature T_w/T_c as expected. Figure 15 shows three examples of how the maximum span-wise temperature ratio T_w^{max}/T_c increases with T_0/T_c. Linear

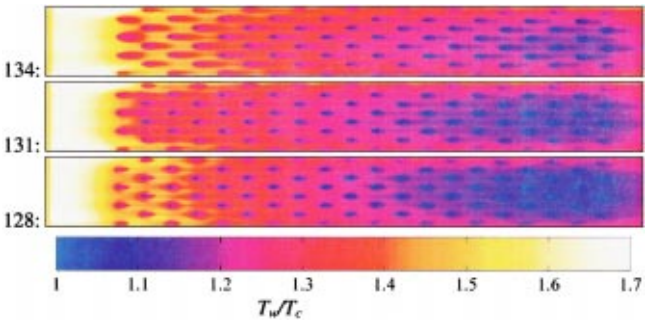


Fig. 7 The effect of different U_0/U_c is shown in thermographs of teflon plates where $T_0/T_c=1.7$, $Re_c=92$, $\delta x/d=6$, $\alpha=20$ deg, and $n=80$. A variation in U_0/U_c is made, from top $U_0/U_c=\{151,100,56\}$.

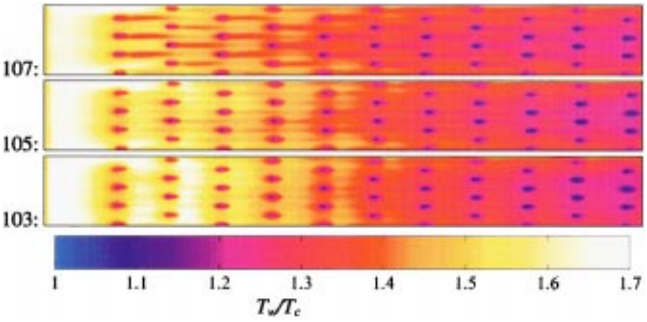
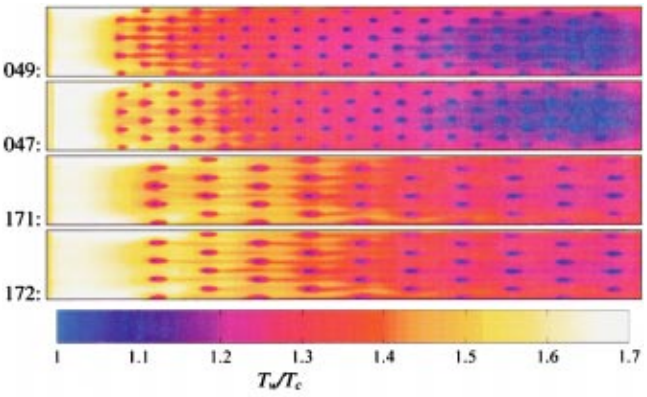
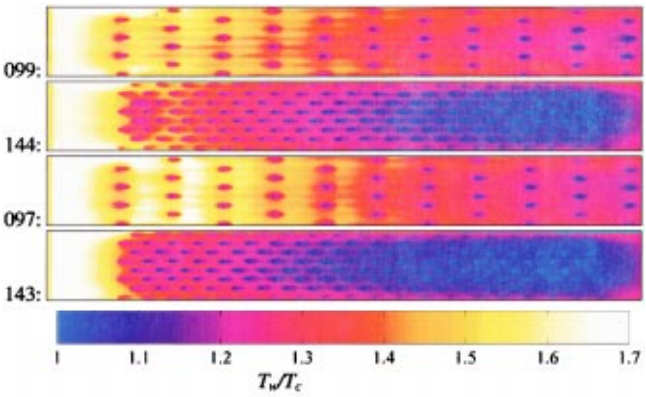


Fig. 8 The effect of different velocity ratios U_0/U_c is shown in thermographs of teflon plates where $T_0/T_c=1.7$, $Re_c=100$, $\delta x/d=12$, $\alpha=20$ deg, and $n=44$. A variation in U_0/U_c is made, from top $U_0/U_c=\{140,90,51\}$.



49: Re_c=153 U_0/U_c=59 $\delta x/d$=6 α=30° n=80
47: Re_c=87 U_0/U_c=59 $\delta x/d$=6 α=30° n=80
171: Re_c=88 U_0/U_c=100 $\delta x/d$=12 α=15° n=40
172: Re_c=138 U_0/U_c=96 $\delta x/d$=12 α=15° n=40

Fig. 9 The effect of different Reynolds numbers Re_c is shown in thermographs of teflon plates with T_0/T_c=1.53–1.59



99: Re_c=97 U_0/U_c=91 $\delta x/d$=12 α=20° n=44
144: Re_c=91 U_0/U_c=102 $\delta x/d$=3 α=20° n=156
97: Re_c=97 U_0/U_c=52 $\delta x/d$=12 α=30° n=44
143: Re_c=91 U_0/U_c=58 $\delta x/d$=3 α=30° n=156

Fig. 11 The effect of different streamwise hole spacings $\delta x/d$ is shown in thermographs of teflon plates where α=20 deg and T_0/T_c=1.56–1.58

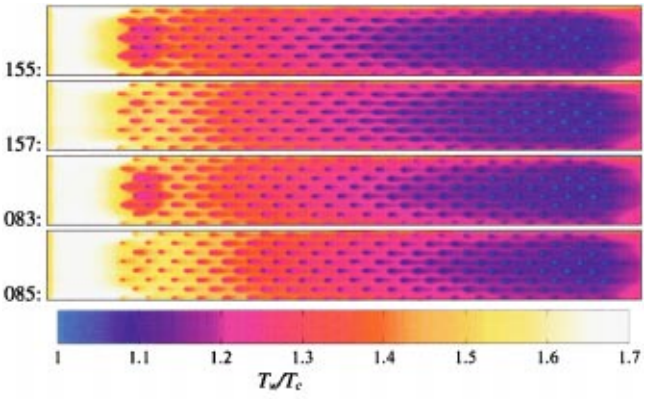
curve fits with error estimates of the predictions (50 percent of the estimated points are within the band) are also plotted. We see that the relationship is approximately true, and can be represented by

$$\frac{T_w^{\max}}{T_c} \approx (1-\eta^{\max})\cdot\left(\frac{T_0}{T_c}-1\right)+1 \tag{7}$$

Rewriting this we see that $\eta^{\max}$ is a film cooling effectiveness parameter. Bazdidi-Tehrani and Andrews [13] reported a weak dependence of T_0/T_c on the film cooling performance. Figure 16 shows that the dependence of T_0/T_c on film cooling effectiveness appears to decrease as U_0/U_c is decreased. Although not shown here, this observation holds for all investigated cases.

When T_0/T_c increases the density ratio ρ_0/ρ_c decreases. This will probably affect the jet liftoff and the relative decrease of enthalpy of the hot crossflow. Figure 16 also shows that the film cooling effectiveness increases as the temperature ratio increases. This effect is more pronounced at high velocity ratios. It should be mentioned that the temperature of the cooling air was nearly constant throughout the experiment, so that the variation in Re_c is due only to a variation in U_c. The effect of temperature dependence on the viscosity and thermal heat conductivity is included in the T_0/T_c parameter.

4.2 Effect of the Velocity Ratio. An increase in the amount of cooling air corresponds to a reduction in the U_0/U_c parameter. This can have a great effect on the surface temperature, as shown



155: Re_c=90 U_0/U_c=102 $\delta x/d$=3 α=20° n=156
157: Re_c=147 U_0/U_c=94 $\delta x/d$=3 α=20° n=156
83: Re_c=88 U_0/U_c=103 $\delta x/d$=3 α=30° n=156
85: Re_c=146 U_0/U_c=94 $\delta x/d$=3 α=30° n=156

Fig. 10 The effect of different Reynolds numbers Re_c is shown in thermographs of teflon plates where T_0/T_c=1.7

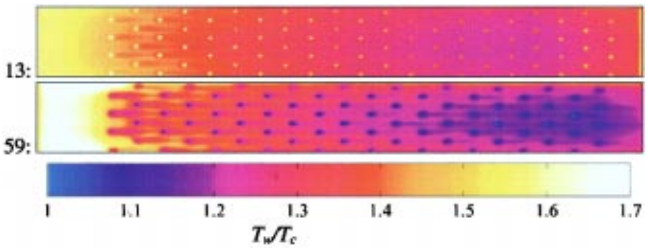
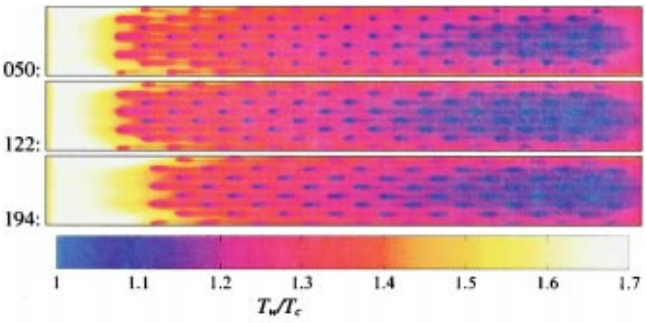


Fig. 12 The effect of different heat conductivity ratios λ_s/λ_c is shown in thermographs of plates with λ_s/λ_c=1700 and n=84 for the steel plate at the top and with λ_s/λ_c=19 and n=80 for the teflon plate at the bottom; furthermore, $\delta x/d$=6, α=30 deg, U_0/U_c=100, and T_0/T_c=1.7



50: Re_c=87 α=30° n=80
122: Re_c=93 α=20° n=80
194: Re_c=88 α=15° n=76

Fig. 13 The effect of different injection angles α is shown in thermographs of teflon plates where U_0/U_c=100, T_0/T_c=1.58, and $\delta x/d$=6

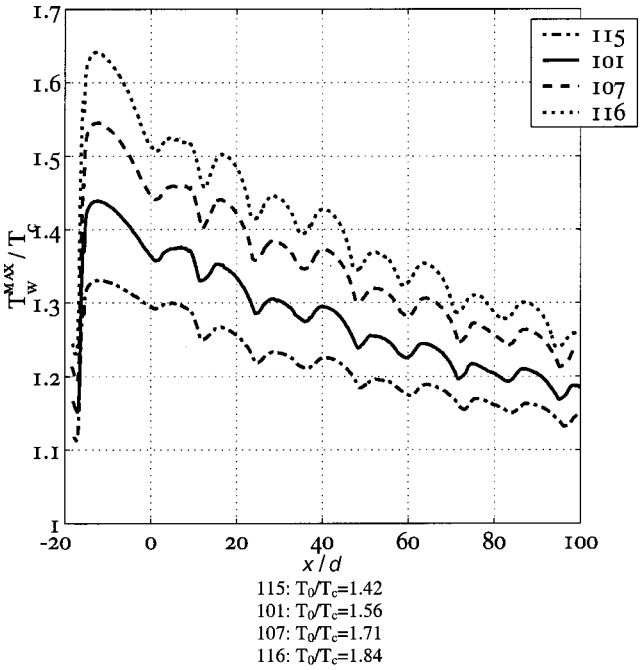


Fig. 14 The effect of T_0/T_c for a teflon plate where U_0/U_c =140, Re_c=100, $\delta x/d$=12, α=20 deg, and n=44

in Fig. 17. In this case the hole spacing is small and thus the jet speed is low (recall that U_c measures the cooling air speed in the plenum chamber and not the velocity of the jet). There is no cooling film for the first rows for the largest U_0/U_c, and the surface temperature is not reduced. The cooling air is distributed only in the downstream holes, probably owing to the pressure drop in the test section; see Fig. 6. This is an important aspect when designing cooling systems fed from a single plenum, as described by, e.g., Salcudean et al. [9]. When the mainstream velocity is reduced, cooling air is distributed in upstream rows, causing an effective decrease in temperature. With a more sparse hole spacing, the jet speed is higher for the same U_0/U_c. The change in cooling effect will not be as pronounced as before; see Figs. 7, 8, 18, and 19. For these cases, the jet penetrates into the mainstream and cooling air is distributed in all holes.

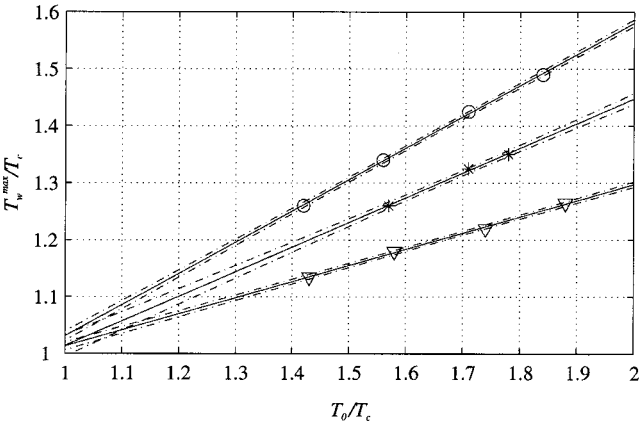


Fig. 15 The maximum spanwise temperature ratio T_w/T_c versus T_0/T_c for three teflon plates at x/d=20. ○: U_0/U_c=140, Re_c=110, $\delta x/d$=12, and α=20 deg; ∗: U_0/U_c=59, Re_c=100, $\delta x/d$=6, and α=30 deg; ▽: U_0/U_c=58, Re_c=90, $\delta x/d$=3, and α=20 deg.

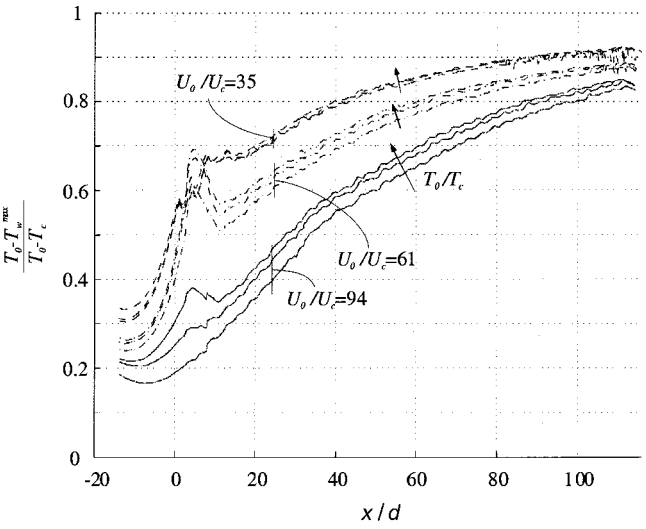


Fig. 16 The film cooling effectiveness η_w^{max} based on maximum spanwise temperature ratio T_w/T_c versus T_0/T_c. Teflon plate with Re_c=148, $\delta x/d$=3, α=30 deg, and n=156.

4.3 Effect of the Reynolds Number. The Reynolds number Re_c has a small effect on the surface temperature at downstream positions on the test plates with $\delta x/d=6$ and $\delta x/d=12$, considering a variation in this parameter of about two times; see Figs. 9 and 20. For the plates with $\delta x/d=3$, there is a significant difference at upstream positions; see Figs. 10 and 21. There seems to be an increase in temperature after the first row of holes in the cases with low Re_c, which is also clearly seen in Figs. 9 and 10. The variation in Re_c is caused only by a variation in U_c, as mentioned in the section covering the T_0/T_c effect.

4.4 Effect of the Streamwise Hole Spacing. The streamwise hole spacing $\delta x/d$ has a major effect on the surface temperature; see Figs. 11 and 22. A dense hole spacing $\delta x/d$ generally gives a low T_w/T_c, if enough cooling air is supplied. The near

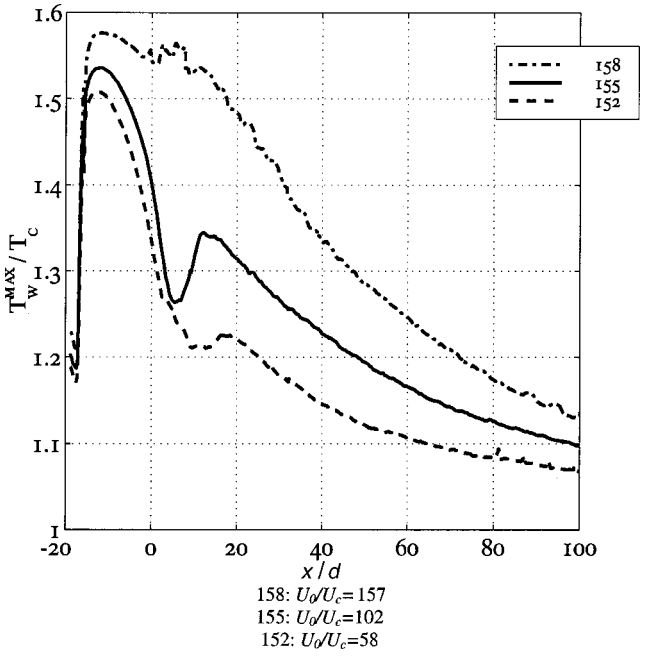
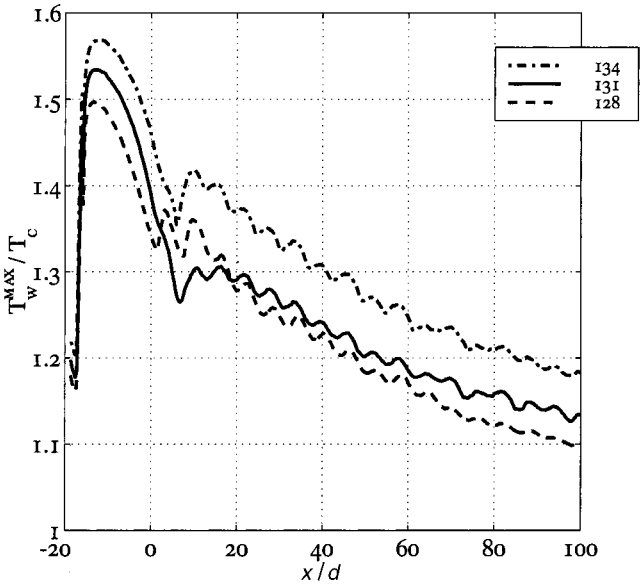
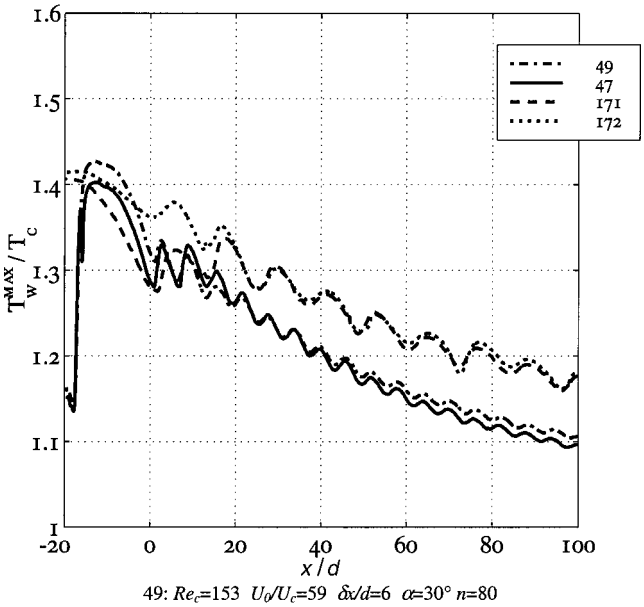


Fig. 17 The effect of U_0/U_c for a teflon plate where T_0/T_c =1.7, Re_c=90, $\delta x/d$=3, α=20 deg, and n=156



134: U_0/U_c=151
131: U_0/U_c=100
128: U_0/U_c=56

Fig. 18 The effect of U_0/U_c for a teflon plate where T_0/T_c =1.7, Re_c=92, $\delta x/d$=6, α=20 deg, and n=80

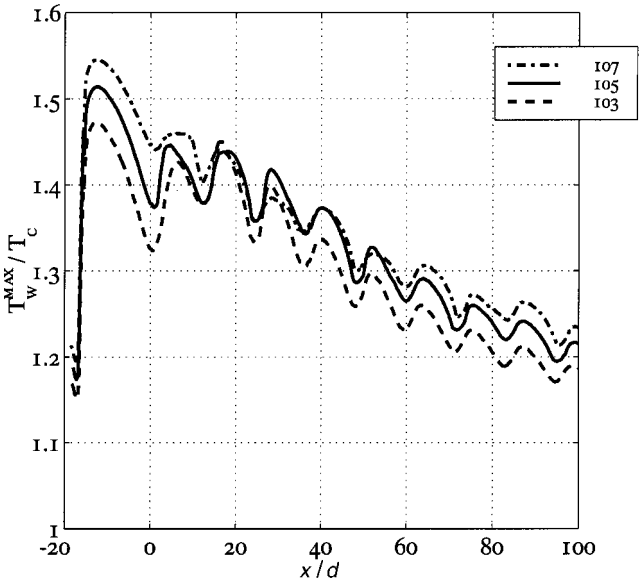


49: Re_c=153 U_0/U_c=59 $\delta x/d$=6 α=30° n=80
47: Re_c=87 U_0/U_c=59 $\delta x/d$=6 α=30° n=80
171: Re_c=88 U_0/U_c=100 $\delta x/d$=12 α=15° n=40
172: Re_c=138 U_0/U_c=96 $\delta x/d$=12 α=15° n=40

Fig. 20 The effect of Re_c for a teflon plate with T_0/T_c =1.53–1.59

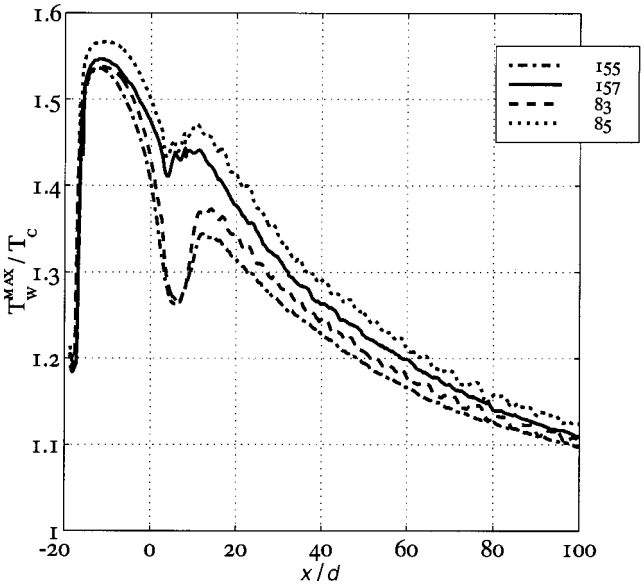
transpiration cooling situation of low $\delta x/d$ seems to drastically reduce the surface temperature, also reported in Bazdidi-Tehrani and Andrews [13]. Higher $\delta x/d$ gives rise to larger oscillations in the temperature profiles in Fig. 22, i.e., larger temperature gradients. For the same U_0/U_c, the plates with dense spacing will have lower jet velocities and more attached jets as compared with the sparsely spaced. The drastic increase in available surface for convective heat transfer inside the wall could also be an explanation for the large effect. Foster and Lampard [5] reported increased film coverage and less jet liftoff when the spanwise hole spacing $\delta z/d$ was decreased.

4.5 Effect of the Thermal Heat Conductivity Ratio. The heat conductivity ratio λ_s/λ_c has an effect on the surface temperature; see Figs. 12 and 23. Higher λ_s/λ_c gives smoother surface temperature profiles, i.e., smaller temperature gradients, as well as a lower temperature at the start of the effusion-cooled plate. A variation in λ_s/λ_c does not seem to have a major influence on either the temperature at the end of the plate or on the spanwise mean temperature level. Note that the plates are



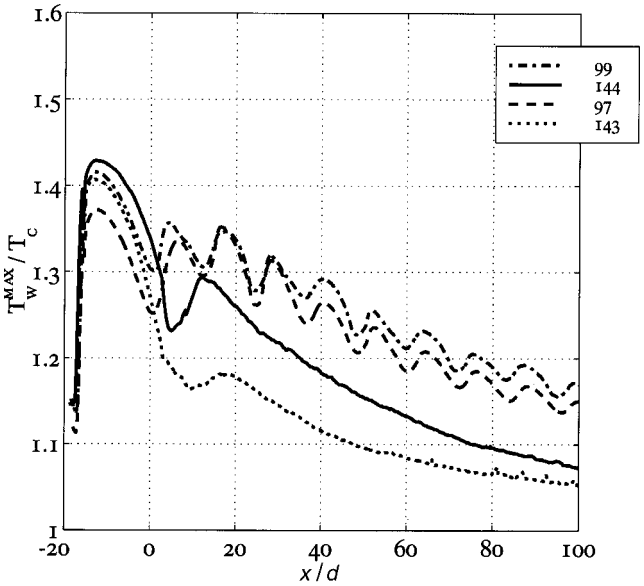
107: U_0/U_c=140
105: U_0/U_c= 90
103: U_0/U_c=51

Fig. 19 The effect of U_0/U_c for a teflon plate where T_0/T_c =1.7, Re_c=110, $\delta x/d$=12, α=20 deg, and n=44



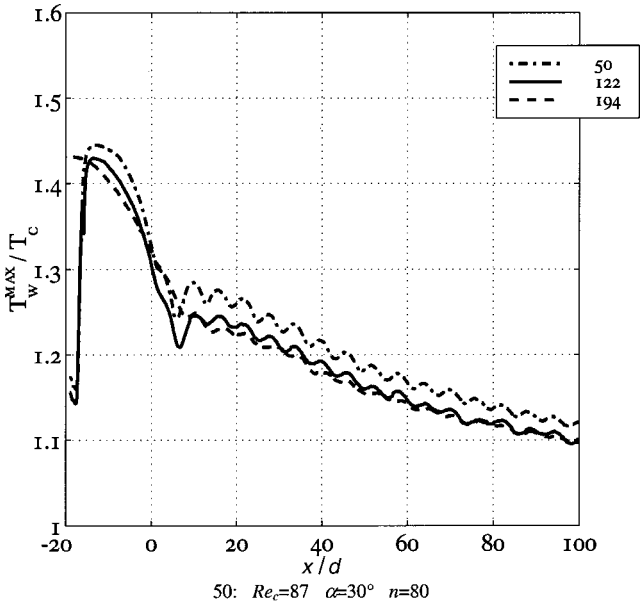
155: Re_c=90 U_0/U_c=102 α=20°
157: Re_c=147 U_0/U_c=94 α=20°
83: Re_c=88 U_0/U_c=103 α=30°
85: Re_c=146 U_0/U_c=94 α=30°

Fig. 21 The effect of Re_c for a teflon plate with T_0/T_c=1.7, $\delta x/d$=3, and n=156



99: Re_c=97 U_0/U_c=91 $\delta x/d$=12 α=20° n=44
144: Re_c=91 U_0/U_c=102 $\delta x/d$=3 α=20° n=156
97: Re_c=97 U_0/U_c=52 $\delta x/d$=12 α=30° n=44
143: Re_c=91 U_0/U_c=58 $\delta x/d$=3 α=30° n=156

Fig. 22 The effect of $\delta x/d$ for a teflon plate with α=20 deg and T_0/T_c=1.56–1.58

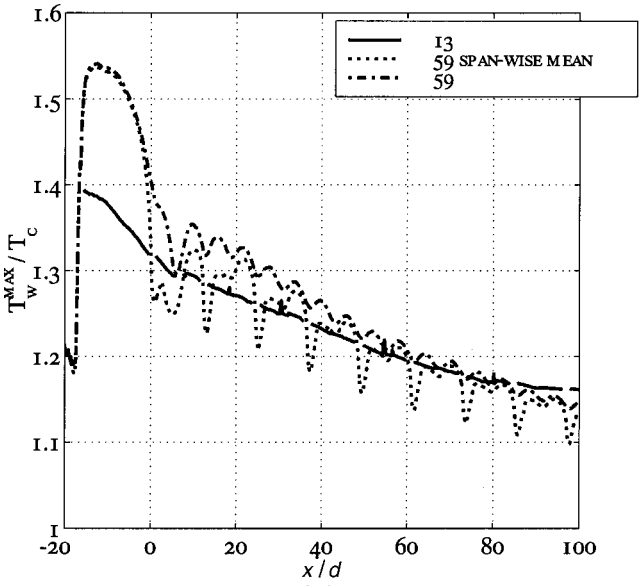


50: Re_c=87 α=30° n=80
122: Re_c=93 α=20° n=80
194: Re_c=88 α=15° n=76

Fig. 24 The effect of α for a teflon plate with T_0/T_c=1.58, U_0/U_c=100, and $\delta x/d$=6

supported at each end and that effects caused by heat conduction may be visible for the steel plate, which has a high thermal conductivity.

4.6 Effect of the Injection Angle. The injection angle α has only a small effect on the surface temperature; see Figs. 13 and 24. The plates with injection angles of 20 and 15 deg are somewhat better than the 30 deg injection hole plates. Shallower injection angles promote attachment of the jet to the wall, but do not seem to lower the overall surface temperature. Low α values reduce the surface temperature of the $\delta x/d=3$ test plates when U_0/U_c is large to a better extent than high α values. It should be mentioned that the lengths of the injection holes are 8, 11.7, and 15.5 diameters for the 30, 20, and 15 deg cases, respectively, if $b/d=4$. It is much more difficult to make a 15 deg injection hole than a 30 deg hole. It has been pointed out by Foster and Lampard [5] and Goldstein and Stone [14] that shallow injection angles give the highest effectiveness for low blowing rates and that steep injection angles perform better at higher blow factors because of increased vortex interaction. From a manufacturing point of view, it is advantageous to keep α as high as possible.



13: Re_c=91 λ_s/λ_c=1700 n=84
59: Re_c=94 λ_s/λ_c=19 n=80

Fig. 23 The effect of λ_s/λ_c for a plate with T_0/T_c=1.7, U_0/U_c=100, $\delta x/d$=6, and α=30 deg

5 Conclusions

A parametric study of 242 cases with different temperature ratios between hot and cold air streams T_0/T_c, velocity ratios U_0/U_c, Reynolds number Re_c, streamwise hole spacing ratio $\delta x/d$, wall material heat conductivity ratio λ_s/λ_c, and injection hole angle α, was performed. Examination of thermographs shows a considerable effect of the conduction in front of the holes due to the blockage caused by the injected jets when U_0/U_c is low. Fully separated jets and hot fluid advected to the wall can also be observed. Even streaks of elevated temperatures behind the injection holes are seen for the low U_0/U_c cases. The surface temperature distribution T_w/T_c seems to increase linearly with the temperature ratio T_0/T_c when the crossflow to jet velocity ratio is low. For dense hole spacing with low jet speeds, the velocity ratio seems to have a large effect, causing uneven distribution of the cooling air. The U_0/U_c parameter seems to decrease in importance as $\delta x/d$ is increased (higher jet speeds and more uniform cooling film distributions). The Reynolds number Re_c has a small effect in downstream positions, but an elevation in temperature was observed for the lowest Re_c after the first row of holes. The closer to the transpiration cooling situation one gets, the better. A small $\delta x/d$ seemed to have a major effect when enough cooling air was supplied. The heat conductivity ratio λ_s/λ_c did not seem to change the temperature in the mean, except in front of the injection holes. The highest temperatures at the upstream part of the plate were reduced. Larger temperature gradients could of course be seen for low λ_s/λ_c. The injection angle did not seem to have a major effect on the surface temperature. Low α gives somewhat lower

temperatures; the 15 and 20° temperature profiles seem to coincide slightly below that of the 30° temperature profiles.

Acknowledgment

The research reported in this paper was supported by the Swedish National Board for Industrial and Technical Development, NUTEK.

Nomenclature

A = area element, [m]
b = thickness of the test plate, [m]
d = diameter of the injection holes, [m]
F_{ij} = view factor between areas A_i and A_j
Ma = Mach number
m = number of surrounding surface elements
n = number of injection holes
Pr = Prandtl number
Re = Reynolds number
T = temperature, [K] or [°C]

Greek letters

α = injection angle
δx = streamwise injection hole spacing, [m]
δz = spanwise injection hole spacing, [m]
Δ = error
ε = diffuse emissivity of the test plate
η = film cooling effectiveness
κ = specific heat ratio (c_p/c_v)
λ = heat conductivity, $[\mathrm{W\,m^{-1}\,K^{-1}}]$
μ = dynamic viscosity, $[\mathrm{kg\,m^{-1}\,s^{-1}}]$
ρ = density, $[\mathrm{kg\,m^{-3}}]$
ρ^D = diffuse reflectivity

Subscripts

0 = the hot main stream
1 = a measurement area on the test plate
amb = ambient condition
c = cooling gas in the plenum chamber
IR = an infrared camera measurement where ε is set to 1.00
jet = the state in the injection holes
s = the solid matrix in the test plate
w = the surface of the test plate

Superscripts

D = diffuse
max = maximum in spanwise direction

References

[1] Gustafsson, K. M. B, 1998, ''An Experimental Study of the Surface Temperature of an Effusion-Cooled Plate using Infrared Thermography,'' thesis for the degree of Licentiate in Engineering, No. 98/9, Department of Thermo and Fluid Dynamics, Chalmers University of Technology, Göteborg, Sweden.

[2] Martiny, M., Schulz, A., and Wittig, S., 1995, ''Full-Coverage Film Cooling Investigations: Adiabatic Wall Temperatures and Flow Visualization,'' ASME International Mechanical Engineering Congress & Exposition San Francisco—November 12–17, Vol. 95-WA/HT-4, ASME, New York.

[3] Cho, H. H., and Goldstein, R. J., 1995, ''Heat (Mass) Transfer and Film Cooling Effectiveness with Injection through Discrete Holes: Part i—Within Holes and on the Back Surface,'' ASME J. Turbomach., **117**, pp. 440–450.

[4] Cho, H. H., and Goldstein, R. J., 1995, ''Heat (Mass) Transfer and Film Cooling Effectiveness with Injection through Discrete Holes: Part ii—On the Exposed Surface,'' ASME J. Turbomach., **117**, pp. 451–460.

[5] Foster, N. W., and Lampard, D., 1980, ''The Flow and Film Cooling Effectiveness Following Injection through a Row of Holes,'' J. Eng. Power, **102**, pp. 584–588.

[6] Friedrichs, S., Hodson, H. P., and Dawes, W. N., 1996, ''Distribution of Film-Cooling Effectiveness on a Turbine Endwall Measured using the Ammonia and Diazo Technique,'' ASME J. Turbomach., **118** No. (4), pp. 613–621.

[7] Gritsch, M., Schulz, A., and Wittig, S., 1998, ''Adiabatic Wall Effectiveness Measurements of Film-Cooling Holes with Expanded Exits,'' ASME J. Turbomach., **120** No. 3, pp. 549–556.

[8] van Treuren, K. W., Wang, Z., Ireland, P. T., and Jones, T. V., 1994, ''Detailed Measurements of Local Heat Transfer Coefficient and Adiabatic Wall Temperature Beneath an Array of Impinging Jets,'' ASME J. Turbomach., **116**, pp. 369–374.

[9] Salcudean, M., Gartshore, I., Zhang, K., and McLean, I., 1994, ''An Experimental Study of Film Cooling Effectiveness near the Leading Edge of a Turbine Blade,'' ASME J. Turbomach., **116**, pp. 71–79.

[10] Ligrani, P. M., and Ramsey, A. E., 1997, ''Film Cooling from Spanwise-Oriented Holes in Two Staggered Rows,'' ASME J. Turbomach., **119**, pp. 562–567.

[11] Eckert, E. R. G., 1984, ''Analysis of Film Cooling and Full Coverage Film Cooling of Gas Turbine Blades,'' ASME J. Eng. Gas Turbines Power, **106** No. 1, pp. 206–213.

[12] Leontiev, A. I., 1999, ''Heat and Mass Transfer Problems for Film Cooling,'' ASME J. Heat Transfer, **121** No. 3, pp. 509–527.

[13] Bazdidi-Tehrani, F., and Andrews, G. E., 1994, ''Full-Coverage Discrete Hole Film Cooling: Investigation of the Effect of Variable Density Ratio,'' ASME J. Eng. Gas Turbines Power, **116**, pp. 587–596.

[14] Goldstein, R. J., and Stone, L. D., 1997, ''Row-of-Hole Film Cooling of Curved Walls at Low Injection Angles,'' ASME J. Turbomach., **119** No. 3, pp. 574–579.

9 ppm NO_x/CO Combustion System for ''F'' Class Industrial Gas Turbines

C. L. Vandervort
Ph.D., P.E.
Manager—Generator DFSS Machine Design,
GE Power Systems,
1 River Road,
Schenectady, NY 12345
e-mail: christian.vandervort@ps.ge.com

The Dry Low NO_x (DLN) -2.6 combustion system has achieved emission rates of lower than 9 ppm NO_x (dry, corrected to 15 percent O_2) and CO from 50 to 100 percent load for the GE MS7001FA industrial gas turbine on natural gas. The system uses lean premixed combustion with fuel staging for low load stability. The first unit achieved commercial operation in March of 1996 with a firing temperature of 2350°F. As of September 9, 1999, it has accumulated over 11,800 hours of operation in peaking and base load service. Sixteen more units have since entered commercial service. Emissions data are shown for operation on natural gas. The DLN-2.6 system can operate on liquid fuel with water injection for NO_x abatement. Power augmentation with steam injection is allowable while operating on natural gas. The premixed gas nozzles utilize swirl for flame stabilization. Aerodynamically shaped natural gas injectors are applied for flashback or flame-holding resistance. [DOI: 10.1115/1.1362661]

Introduction

The regulatory requirements for low emissions from gas turbine power plants have become increasingly more stringent during the past 15 years. Environmental agencies throughout the world are now requiring even lower emissions of NO_x and other pollutants from both new and existing gas turbines. Traditional methods of reducing NO_x emissions from combustion turbines (water and steam injection) are limited in their ability to reach the extremely low levels required in many localities. Dry Low NO_x (DLN) combustors control emissions by use of excess air as a diluent in conjunction with careful control of flame stoichiometry. Early industrial applications of DLN technology have been described by Washam [1], and Davis and Washam [2].

A new DLN combustion system, the ''DLN-2.6,'' has entered commercial operation with emission levels of less than 9 ppm NO_x corrected to 15 percent O_2 and less than 9 ppm CO from approximately 50 to 100 percent load at ''F'' class conditions while operating on natural gas. These results were achieved on a GB MS7001FA (7FA) with a firing temperature of 2350°F. The DLN-2.6 combustor evolved from the DLN-2 system that operates at 25/15 ppm NO_x/CO emissions described by Davis [3]. In addition to achieving single-digit emissions and extended turndown, the DLN-2.6 improves low load CO and unburned hydrocarbon (UHC) emissions relative to the DLN-2. This new system can be retrofitted to existing 7FA DLN-2 gas turbines.

9 ppm NO_x Development

DLN systems achieve low NO_x levels by operating at very lean conditions. NO_x production is highly dependent upon flame or reaction zone temperature as proposed by Zeldovich [4]. The excess air dilutes the flame zone in a manner that suppresses peak reaction zone temperatures and minimizes NO_x production. Reduction of NO_x levels from the DLN-2 at 25 ppm to 9 ppm required that approximately 6 percent additional air needed to pass through the premixers. This change was accomplished through reductions in cap and liner-cooling flows that necessitated improved cooling effectiveness. Lower flame temperatures reduced NO_x production rates. However, without changes in staging, turndown would have been adversely impacted.

The key feature of the new configuration is the addition of a sixth burner located in the center of the five existing DLN-2 burners. The presence of the center nozzle enables the DLN-2.6 to extend its turndown well beyond the five-burner DLN-2. By fueling the center nozzle separately from the outer nozzles, the fuel/air ratio can be modulated relative to the outer nozzles. The fuel/air ratio in the center burner can be reduced below its local lean flammability limit while maintaining the outer burners above their lean limit. In this way the overall combustor fuel/air ratio can be reduced to a lower value than would otherwise be possible with uniformly fueled burners.

A new staging scheme was proposed to start the machine and operate at low load in premixed mode. Otherwise, addition of this center nozzle to the DLN-2 system would have required five fuel manifolds for the DLN-2.6 compared to four on the DLN-2. The new configuration uses four circuits with three premixed manifolds staging fuel to the six burners, plus a fourth premixed manifold, quaternary fuel, for dynamics abatement. The first three premixed manifolds, designated PM1, PM2, and PM3, are configured such that any number (1 to 6) of burners can be operated at any time. The PM1 manifold fuels the center nozzle, the PM2 manifold fuels two outer nozzles, and the PM3 manifold fuels the remaining three outer nozzles.

Development of the 9 ppm NO_x system was accelerated through application of DLN-2 technology and hardware. The outer fuel nozzles are identical to those used by the DLN-2 system, where they have been proven to provide excellent premixed flame stability and strong premixing performance over the necessary operating range. The liner cooling method evolved from the DLN-2 configuration that uses enhanced convection of the outer surface. Both the DLN-2 and DLN-2.6 incorporate an impingement cooled transition piece.

Mechanical Configuration

The DLN-2.6 combustion system is a single-stage combustor that can operate on either gaseous or liquid fuel. Configuration of the DLN-2.6 combustion system is of the ''can-annular'' type shown by Fig. 1.

Compressor discharge air exits a diffuser to enter the compressor discharge casing where it supplies impingement cooling to the transition piece. Flow moves forward through an annular space

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Munich, Germany, May 8–11, 2000; Paper 00-GT-086. Manuscript received by IGTI Feb. 2000; final revision received by ASME Headquarters Jan. 2001. Associate Editor: M. Magnolet.

created between the flow sleeve and liner. A small percentage is supplied for cap cooling prior to collecting in the region upstream of the fuel nozzles. From this region, air moves through the swirlers and past the premixed gas fuel injection ''pegs'' into a premixing zone. Flame ''anchors'' at a point defined by the aft end of the fuel nozzle and the cap face. The flame is contained by the liner, and the resulting products of combustion are directed into the transition piece for entry into the first-stage turbine nozzle annulus.

Figure 2 is a schematic of the fuel nozzle configuration. The five nozzles located around the circumference of the end cover are referred to as outer nozzles and the single (PM1) nozzle is the center nozzle. ''PM'' refers to premixing of the fuel and air prior to combustion, and the number following PM refers to the number of nozzles per chamber supplied by each manifold.

The quaternary fuel passage differs from the others in that it is located upstream of the premixers. It consists of fuel pegs as opposed to nozzles. These approximately 2-in.-long pegs are spaced circumferentially around the casing. This fuel is delivered upstream of the fuel nozzles in small proportions of total fuel flow and only at high loads for control of combustion dynamics is the only function of quaternary fuel. Dynamics are also suppressed through ''tuning'' of the PM1, PM2, and PM3 fuel splits.

Description of Major Components

Fuel Nozzles. Figure 3 shows a cross section of a DLN-2 or 2.6 outer gas fuel nozzle. The nozzle has passages for diffusion and premixed gas. Diffusion passages for the DLN-2.6 are supplied a continuous air purge internally through the end cover. Premixed passages of two of the five outer nozzles are fed from the PM2 manifold. Fuel oil flows through the center of the cartridge to a nozzle at the tip. Water for NO_x abatement flows in an annular passage surrounding the oil tube. Atomizing air is supplied from off-base compressors via the end cover to the annular passage formed by the outer diameter of the water tube and the inner surface of the gas nozzle. The cartridge is breech loaded through the end cover with external connections. Gas-only units apply ''blank'' cartridges of similar construction but without the oil and water passages. Air purge to the blanks is also supplied through the end cover internal passages.

The center fuel nozzle is of similar configuration but with a simplified geometry. The swirler is slightly smaller, and the diffusion, oil, water, and atomizing air passages are eliminated. A ''passive'' purge flow is supplied from upstream of the swirler and directed to the tip of the fuel nozzle.

Fuel nozzles shipped for the original DLN-2.6 system were improved by incorporation of a flashback ''fuse.'' Flashback occurs when flame can ''hold'' or be supported in the recirculation zone downstream of the premixed gas pegs. In the event of a flashback, the ''fuse'' on the nozzle tips opens, resulting in diversion of gas fuel from the premixed pegs to a point further downstream. The fuel will enter the airflow and support a flame at the tip of the nozzles, similar to that established in diffusion mode for the DLN-2. Remaining fuel to the premixed pegs is insufficient to support a flame in that region, and the flashback condition is corrected. The unit will continue to operate with a very slight change in exhaust temperature profile and NO_x. Machine reliability would not be impacted. In general, an open fuse event for a single fuel nozzle would not be detected by a continuous emissions monitoring system (CEMS). The affected nozzle would be identified and removed for repair during a regularly scheduled combustion inspection.

In addition to the nozzle fuse, design activities were initiated to preclude flame holding on the premixed pegs. The solution was to add full ''fairings'' to the downstream surface of the cylindrical fuel injection pegs. Partial fairings were originally applied to the DLN-2.6 system as an upgrade from the DLN-2 fuel nozzle. Laboratory testing demonstrated that full fairings were even more effective in reducing the probability of fuel nozzle flashback. Fully faired nozzles were also applied to the DLN-2 fleet where the design has accumulated over 500,000 fired hours of experience as of September 1999 without a single flashback event.

End Cover Assembly. The end cover functions as the interface between the fuel and purge piping to the fuel nozzles, and provides the pressure barrier for the ''head'' end of the combus-

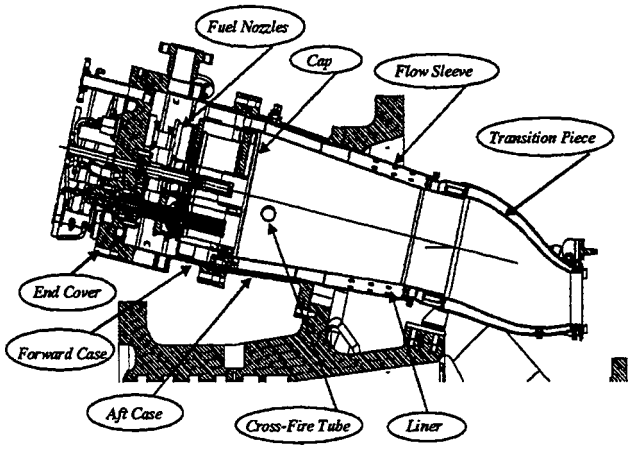


Fig. 1 DLN-2.6 cross section

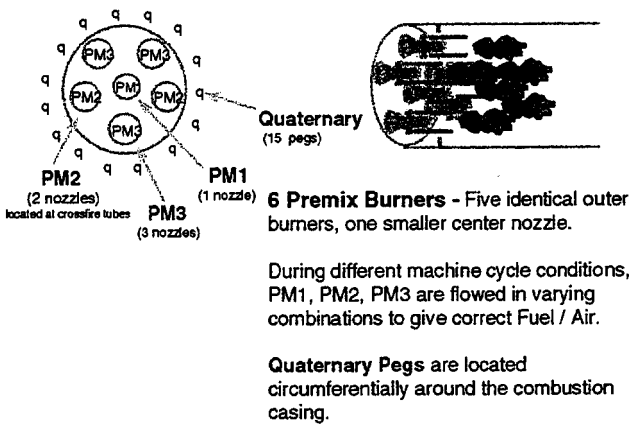


Fig. 2 DLN-2.6 fuel nozzle arrangement

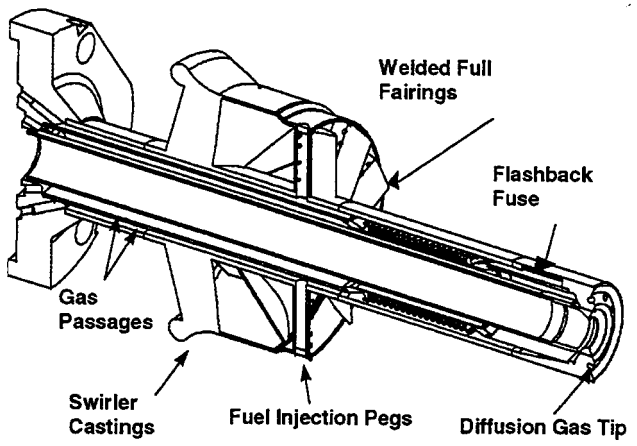


Fig. 3 DLN-2.6 outer fuel nozzle cut-away diagram

tion system. It is a fabrication with internal fuel manifold ''rings'' and brazed insert passages leading to each fuel nozzle. Long neck flange connections are welded to the outer surface to connect with the fuel gas supply ''pigtail'' piping. The gas fuel nozzles are bolted to the inner surface with gaskets to prevent leakage. Dual fuel units have the breech loaded oil and water cartridges versus the gas-only ''blank'' cartridges.

Liner. The liner provides the mechanical barrier between the flame zone and other components of the combustion system. It is cylindrical with a taper from the ''head'' end to the entrance of the transition piece. Repeated ribs (turbulators) are utilized on the air side to provide enhanced cooling. Thermal barrier coating is applied to the gas side. In order to minimize film cooling and aft seal leakage, a new aft end configuration referred to as ''2-COOL'' was developed. 2-COOL is a double-walled cylindrical section having multiple axial cooling channels around its circumference. The high aspect ratio of the cooling passages provides very high heat transfer coefficients with small amounts of cooling flow. Air feeds into the channels from the flow sleeve annulus, flows through the cooling channels, and dumps to the downstream edge of the liner into the transition piece. Reference [5] provides a detailed description of the 2-COOL system.

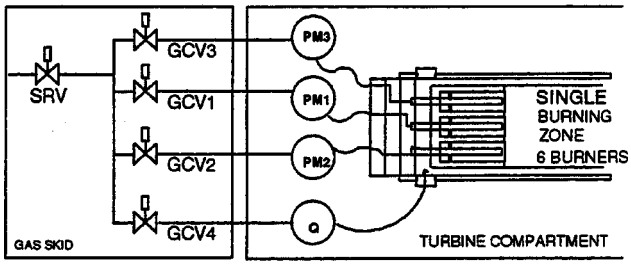
Cap Assembly. The cap assembly provides the upstream barrier between the flame zone and the fuel nozzle and cover assembly. The DLN-2.6 applies a structure that is very similar to that in operation for the DLN-2's. The difference is that cap face cooling was redesigned to minimize cooling air consumption. The DLN-2.6 moved to ''effusion'' cooling with angled cooling holes for improved cooling effectiveness.

Flow Sleeve. The flow sleeve provides the outer radial surface for the annular passage surrounding the liner. Combustion air flows through this passage toward the end cover plenum for supply to the fuel nozzle air swirlers. Impingement holes are placed near the aft (entry) end of the flow sleeve to provide additional cooling to the aft regions of the liner.

Casings. DLN-2.6 uses both forward and aft casings. These items combine with the end cover to provide the structural pressure barrier. The forward casing contains the cap and receives the bolting from the end cover and fuel nozzle assemblies. The combination of the forward casing, end covers, fuel nozzles, and cap is often installed and removed as a single major assembly. The aft casing encircles the flow sleeve and liner. Cross-fire tubes are positioned between adjacent liners that penetrate the aft casings. Four of the upper chambers incorporate flame detectors. Two other upper chambers have spark plugs. The spark plugs are fully retractable. They are forced outward by chamber pressure on unit acceleration, and reinserted on shut down by spring force. A dynamic pressure probe is located on each chamber for monitoring of combustion dynamics. This port is also useful as a boroscope entry point.

Transition Piece Assembly. The DLN-2.6 transition piece assembly is very similar to the DLN-2 or multi-nozzle quiet combustor (MNQO) design which has been in service for nearly ten years. The transition piece is a duct that accepts the combustion gases from the liner and directs their flow into the turbine annulus. Cooling for all ''F'' class transition pieces is provided by impingement through an outer sleeve. The ''aft'' frame of the transition piece receives convection and film cooling from the flow of air through numerous small cooling passages. Floating metallic seals connect the transition piece assembly to the first stage nozzle ring. Metallic ''cloth'' seals have recently been added to the product offering for improved sealing capability, reduced system wear, and greater reliability.

Controls. The integrated DLN system consists of a staged premixed combustor, the gas turbine's controls, and the fuel systems. The controls must be capable of meeting required emission levels on both gas and oil fuel. The system must also provide for operability with emphasis on smooth and reliable combustor mode changes, ability to load and unload, capability to switch from one fuel to another, and system response to rapid transients (e.g., generator breaker open events or rapid swings in load). The design goal is to make the controls system function such that the presence of the DLN system is transparent to the operator.



SRV SPEED/RATIO VALVE
GCV1 GAS CONTROL PM1
GCV2 GAS CONTROL PM2
GCV3 GAS CONTROL PM3
GCV4 GAS CONTROL Quaternary

PM3 - 3 NOZ. PRE-MIX ONLY
PM2 - 2 NOZ. PRE-MIX ONLY
PM1 - 1 NOZ. PRE-MIX ONLY
Q - QUAT MANIFOLD, CASING, PRE-MIX ONLY

Fig. 4 DLN-2.6 fuel distribution and controls system

The DLN-2.6 control system regulates the fuel distribution to the PM1, PM2, PM3, and quaternary fuel system. The fuel flow distribution to each combustion fuel system is a function of combustion reference temperature and inlet guide vane (IGV) position. Modes 1, 2, 3, 4, 5Q, and 6Q are established by changing the distribution of fuel flow in the combustor. The gas fuel system shown by Fig. 4 consists of the gas fuel stop/ratio valve, PM1 gas control valve, PM2 gas control valve, PM3 gas control valve, and quaternary gas control valve. The stop/ratio valve is designed to maintain a predetermined pressure at the control valve inlet. The PM1, PM2, PM3, and quaternary gas control valves regulate the desired gas fuel flow delivered to the turbine.

Fuel Staging Strategy. Fuel to the DLN-2.6 combustor is staged to operate the machine over the entire load range. Burners are brought on in stages, starting at full speed no load (FSNL) with the center burner only and turning on additional burners as load is increased. The staging is accomplished by using four fuel systems. Each fuel system consists of a different number of nozzles coupled together via a common manifold. These systems are used alone or in combination with one another to maintain fuel/air ratio within a desirable range in the reaction zone. The mode names reflect the number of burners that are being fueled in that mode. The suffix Q in modes 5Q and 6Q reflects the fact that quaternary fuel may be in use. Mode 5Q and 6Q combine to span the load range for 9/9 ppm NO_x/CO operation.

The machine lights on the PM2 and PM1 circuits. Following ignition, the PM1 nozzles are turned off and the PM2 nozzles support combustion while the unit ramps to 95 percent speed. At 95 percent speed, the unit transfers from mode 2 to mode 1 for continued acceleration to full speed no load and leading to synchronization. At approximately 10 percent load, the PM2 manifold is turned back on (3 nozzles=mode 3) and continues until 25 percent load. At this transfer point, the PM2 manifold is switched off, with the PM3 manifold turned on for operation in mode 4 to 40 percent load. Mode 5 consists of the outer nozzles only; quaternary fuel may be used to become mode 5Q. Finally, at 45 percent load, the center nozzle is turned on to enter mode 6 or 6Q with quaternary. Figure 5 summarizes the ignition, cross-fire, acceleration, and loading sequence.

The unloading sequence is an approximate reversal as shown in Fig. 6, but with the transfer ''out'' temperatures reduced by about 60°F. This provides the maximum load turndown conditions. For

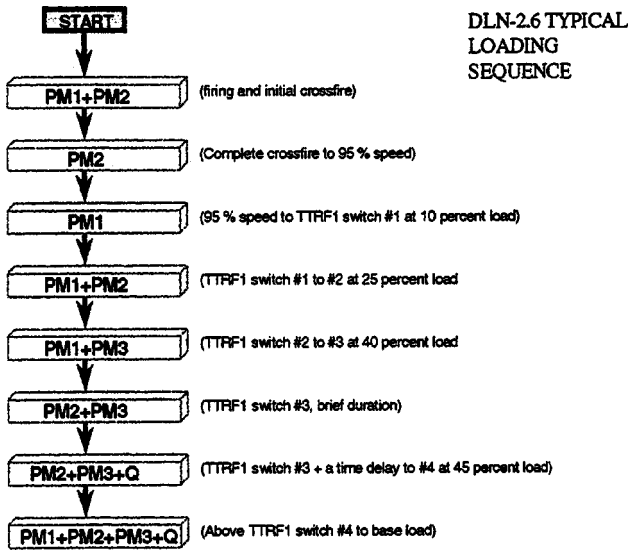


Fig. 5 DLN-2.6 ignition, cross-fire, acceleration, and loading strategy

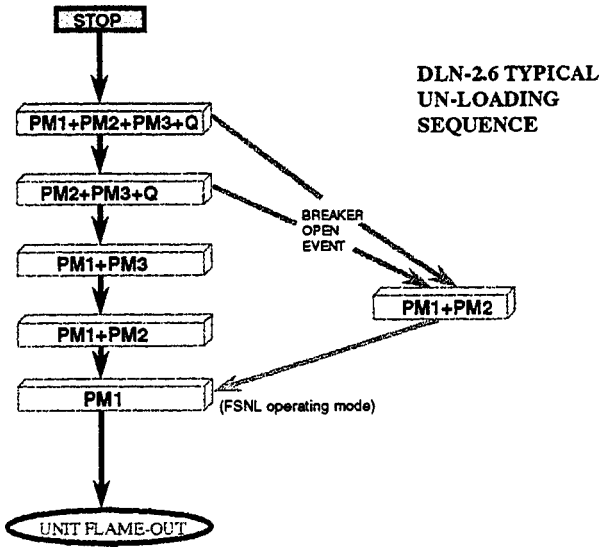


Fig. 6 DLN-2.6 unloading and fired shutdown sequence

example, the mode 5 to 4 transfer occurs at lower than 30 percent load. On a breaker open event, the load rejection to FSNL uses the PM2 to support the flame during the early stages of the transient, but then shifts over to the PM1 in preparation for steady state operation at FSNL. On a fired shutdown, the unit will decelerate in mode 1 to flame-out, which occurs at approximately 25 percent speed.

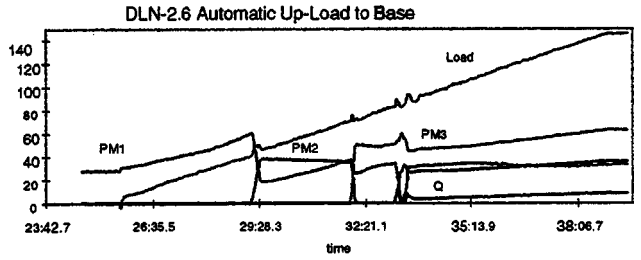


Fig. 7 DLN-2.6 loading transient from FSNL to base load

Figure 7 shows a representative loading transient from FSNL to base load as a function of time. Actual load in megawatts is shown along with the relative fuel flows in each of the PM1, PM2, PM3, and quaternary fuel circuits. In this case, loading from FSNL to base required approximately 15 minutes. Load transients during mode changes were of acceptable amplitudes and time duration.

DLN-2.6 Combustion System Performance

Field Test Results on Natural Gas. The first DLN-2.6 unit was fully capable of functioning over the entire load range with no hardware changes, and was operated in ''peaking'' service throughout 1996. From December of 1996 through March of 1997, changes were made to the end cover and cap designs to improve mechanical robustness. Over this time interval, emissions were consistently below 9 ppm for both NO_x and CO from 50 to 100 percent load. Conversion to the flashback resistant fuel nozzles in February of 1998 caused a temporary increase in NO_x levels that was corrected seven months later. Since that time, NO_x and CO levels have, again, been below 9 ppm. Figures 8 and 9 show the emissions performance for natural gas fuel. Combustion dynamics (noise) data were collected for each chamber at all load conditions. Measured values were below the general limits for continuous operation in order to meet reliability and hardware durability objectives.

Distillate Fuel Performance with Water Injection—Full Pressure Laboratory Testing. The 7FA DLN-2.6 system was designed to achieve emissions of 42 ppm NO_x corrected to 15 percent O_2 and 20 ppm CO on fuel oil. Proven hardware components from the Frame 7FA and 9FA DLN-2 program were applied directly to this system. Each of these configurations achieved their emissions goals in commercial applications. Single combustion chamber laboratory testing was performed with the DLN-2.6 configuration, with results shown by Fig. 10. As of March of 1999, all DLN-2.6 units commissioned have been gas-only systems. Therefore, field test results are not yet available.

Steam Injection for Power Augmentation on Natural Gas Injection—Full Pressure Laboratory Testing. Industrial gas turbine customers routinely request steam injection capability for

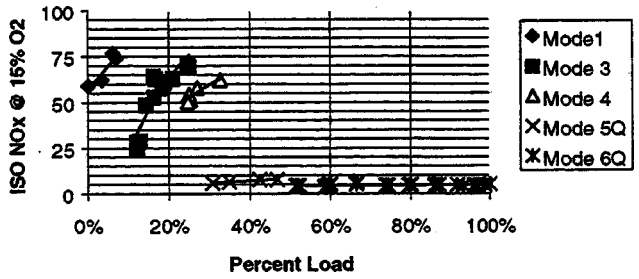


Fig. 8 NO_x at 15 percent O_2 versus percent load

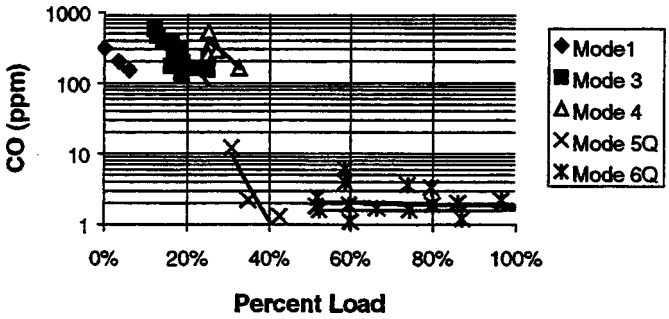


Fig. 9 CO levels versus percent load

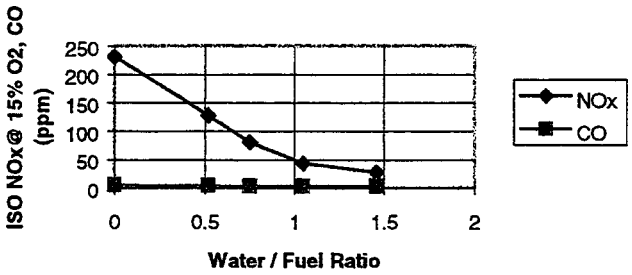


Fig. 10 Base load equivalent emission performance on distillate fuel with water injection

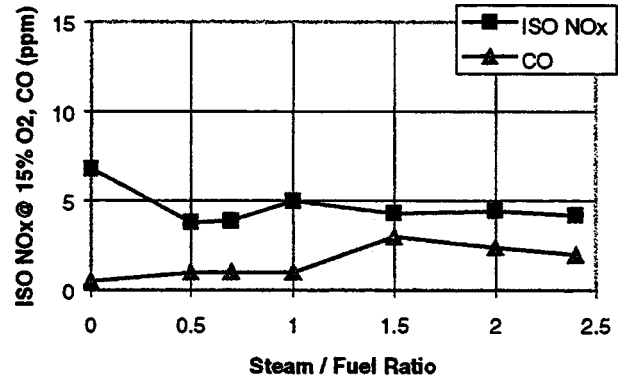


Fig. 11 Base load emission performance with steam injection for power augmentation on natural gas

power augmentation on natural gas. The first two Frame 7FA DLN-2.6 units with steam injection capability achieved commercial operation this year. Steam for these units enters through the forward casing into the plenum established by the end cover and cap region. The steam mixes with combustion air as it passes through the nozzle swirlers prior to introduction of fuel gas and entry into the premixing zone.

Laboratory testing with steam injection was performed with excellent results, as illustrated by Fig. 11. The presence of the center fuel nozzle is very important to maximizing the total steam injection flow. Fuel/air ratio can be optimized by increasing or decreasing the center fuel nozzle flow to establish the lowest possible emissions with good flame stability and low combustion dynamics.

Operation and Maintenance

Hoeft and Gebhardt [6] provide a complete description of operation and maintenance practices for industrial gas turbines. The DLN-2.6 combustor was designed for operation in many varied applications. The first unit has been employed primarily as a ''peaker'' with daily startups and shutdowns operating as a simple cycle gas turbine. During mid-1998, the station erected a heat recovery steam generator (HRSG) for coupling to an existing steam turbine for combined cycle service. The second 7FA DLN-2.6 to reach commercial operation is arranged as a single-shaft ''STAG'' combined cycle that operates as a base load station. As of September 9, 1999, this second unit had accumulated over 13,000 hours with 64 fired starts.

Combustion inspections are required at regular intervals for all systems. The DLN-2.6 meets an 8000 hour/400 start interval for operation on natural gas. A development program is underway with the goal of reaching a 12,000 hour inspection interval. Maintenance ''acceleration'' factors are historically applied for operation under any of the following conditions:

Table 1 Estimated DLN-2.6 repair and replacement cycles

"CI" = Combustion Inspection	Repair Interval	Replace Interval (Hours)	Replace Interval (Starts)
Combustion Liners	CI	5 (CI) *	5 (CI) *
Transition Pieces	CI	5 (CI) *	5 (CI) *
Fuel Nozzle/Cross Fire Tubes	CI	3 (CI)	3 (CI)

* For current hardware, the maximum anticipated replacement interval is given with a goal of 40,000 hours.

- Steam injection for power augmentation on natural gas
- Operation at ''peak'' load in excess of the rated base load
- Distillate fuel
- Distillate fuel with water injection for NO_x abatement

Operation under any of these circumstances will reduce the recommended interval between combustion inspections.

The combustion inspection for a DLN-2.6 generally requires four days assuming a crew size of six working 12 eight-hour shifts. This assumes that replacements for all parts in need of repair are available on site. Repair and replacement of combustion hardware must be anticipated over the life of the unit, as estimated by Table 1.

Conclusion

Dry low NO_x programs continue to focus on achieving the extremely low NO_x levels required to meet today's regulations, and to prepare for more stringent requirements in the future. DLN-2.6 machines operating at F technology levels are routinely operating below 9 ppm NO_x and CO from 50 to 100 percent load. Laboratory testing has been successfully completed, which demonstrates the capability to operate on distillate fuels with water injection to NO_x levels of less than 42 ppm. Steam injection capability for power augmentation while operating on natural gas has also been demonstrated. Combustion dynamics are sufficiently low in all cases for the system to meet reliability and hardware life requirements. Future development is focused upon reaching a combustion inspection interval of 12,000 hours, surpassing the industry standard of 8000 hours. The system is being applied to all 7FA units requiring emission levels of 15 ppm or less. In total, 16 units were operating commercially as of September 1999.

Acknowledgments

The DLN-2.6 combustion system has achieved success through the dedicated efforts of many talented engineers working over the span of the last several years. Those providing the most significant contributions include Dr. Warren Mick, Mr. Charles Steber, Dr. L. Berkley Davis, Mr. David Fitts, Mr. John Cole, Ms. Johanna Wellington, Ms. Kristina Cairns, Mr. Richard Bourgeois, and Ms. Leanne Peduzzi.

References

[1] Washam, R. M., 1983, ''Dry Low NO_x Combustion System for Utility Gas Turbine,'' ASME Paper 83-JPGC-GT-13.
[2] Davis, L. B., and Washam, R. M., 1989, ''Development of a Dry Low NO_x Combustor,'' ASME Paper No. 89-GT-255.
[3] Davis, L. B., 1997, ''Dry Low NO_x Combustion System for GE Heavy-Duty Gas Turbines,'' GE Power Systems Reference GER-3568F.
[4] Zeldovich, J., 1946, ''The Oxidation of Nitrogen in Combustion and Explosions,'' Acta Physicochim. URSS, **21**, No. 4, pp. 577–628.
[5] Ritter, A. M., et al., 1998, ''Combustor for a Gas Turbine with Cooling Structure,'' United States Patent Number 5,724,186, March 10.
[6] Hoeft, R. F., and Gebhardt, E., 1998, ''Heavy-Duty Gas Turbine Operating and Maintenance Considerations,'' GE Power Systems Reference GER-3620F.

Computational Modeling of Self-Excited Combustion Instabilities

S. J. Brookes
R. S. Cant[1]
I. D. J. Dupere
A. P. Dowling

CFD Laboratory,
Department of Engineering,
University of Cambridge,
Trumpington Street,
Cambridge CB2 1PZ, United Kingdom
email: rsc10@eng.cam.ac.uk

It is well known that lean premixed combustion systems potentially offer better emissions performance than conventional non-premixed designs. However, premixed combustion systems are more susceptible to combustion instabilities than non-premixed systems. Combustion instabilities (large-scale oscillations in heat release and pressure) have a deleterious effect on equipment, and also tend to decrease combustion efficiency. Designing out combustion instabilities is a difficult process and, particularly if many large-scale experiments are required, also very costly. Computational fluid dynamics (CFD) is now an established design tool in many areas of gas turbine design. However, its accuracy in the prediction of combustion instabilities is not yet proven. Unsteady heat release will generally be coupled to unsteady flow conditions within the combustor. In principle, computational fluid dynamics should be capable of modeling this coupled process. The present work assesses the ability of CFD to model self-excited combustion instabilities occurring within a model combustor. The accuracy of CFD in predicting both the onset and the nature of the instability is reported. [DOI: 10.1115/1.1362662]

Introduction

Increasing legislative pressure on pollutant emissions has led to the adoption of lean premixed combustion technology for industrial gas turbines. Lean premixed operation offers several advantages, including lower combustion temperatures leading to a drastic reduction in thermal NO_x, accompanied by lower emissions of soot and CO. Nevertheless there are some serious difficulties resulting from reduced flame stability and increased susceptibility to external disturbances such as acoustic perturbations. Even moderate coupling between heat release and acoustic pressure fluctuations can lead to self-excited combustion instability accompanied by unacceptable levels of noise, while stronger coupling can result in large pressure excursions and possible structural damage. Clearly it is vital that the combustion system designer has the information necessary to ensure that such instabilities can be avoided or controlled. Traditionally, such information has been obtained from experiment, but there is growing evidence that techniques based on computational fluid dynamics (CFD) are now sufficiently reliable to be able to make a valuable contribution to the design process.

Computation of the unsteady flow and turbulent premixed combustion in a gas turbine combustion chamber poses a serious challenge to current CFD techniques, and places severe demands on both the physical modeling and the numerical analysis. It is necessary to be able to predict the acoustic pressure field and the rates of heat release within the flame to a high degree of accuracy in order to capture the frequency and magnitude of the coupled oscillations. For maximum value in terms of combustion system design it is also important to be able to predict the onset of unstable combustion, and this requires even greater sensitivity from the modeling. Previous work in this area has been able to capture fully developed combustion-driven oscillations using the unsteady Reynolds-averaged Navier-Stokes equations ([1,2]) and good agreement with experimental data has been obtained ([3]). The inherent unsteadiness of the problem makes it a good candidate for large eddy simulation using a variety of combustion submodels ([4,5]), and successful comparisons with experiment have been made ([6]). Theoretical methods have been able to predict stability boundaries ([7]), but to date there has been no validated CFD prediction of the onset of self-excited combustion instability.

The aim of the present work is to investigate the use of CFD for the prediction of self-excited combustion instability in lean premixed systems. The work builds directly upon that of Brookes, Cant, and Dowling [3] which demonstrated the value of CFD in predicting forced oscillations. In the forced case the amplitude and frequency of the oscillation are imposed externally and the response of the system is measured. The self-excited case is more challenging in that no oscillation is imposed externally and hence all unsteadiness must originate from the solution itself. The test case is the experiment of Langhorne [8] in which a premixed flame is stabilized on a bluff centerbody in a long axisymmetric duct. The experimental results for a broad range of fuel-air equivalence ratios indicated the existence of two distinct types of self-excited oscillation, which were named ''weak buzz'' and ''established buzz.'' Both exhibited coupling between the turbulent flame and the acoustically active pressure field, but the established oscillation which occurred at higher equivalence ratio was observed to have higher frequency as well as higher amplitude than the weak oscillation.

In order to try to capture this behavior, CFD simulations were carried out over a range of fuel-air mixture strengths and the results for the unsteady reacting flow field were compared with pressure, velocity, and heat release data from the experiment. It was found possible to predict both types of self-excited behavior, and most importantly to predict the transition between them, all in good agreement with the experimental results.

Formulation

The governing equations for the present work are the compressible Navier-Stokes equations for mass, momentum, and energy conservation, augmented by a transport equation for a reaction progress variable defined as

$$c = \frac{Y_{FR} - Y_F}{Y_{FR} - Y_{FP}}. \quad (1)$$

[1]Corresponding author.

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Munich, Germany, May 8–11, 2000; Paper 00-GT-104. Manuscript received by IGTI February 2000; final revision received by ASME Headquarters January 2001. Associate Editor: M. Magnolet.

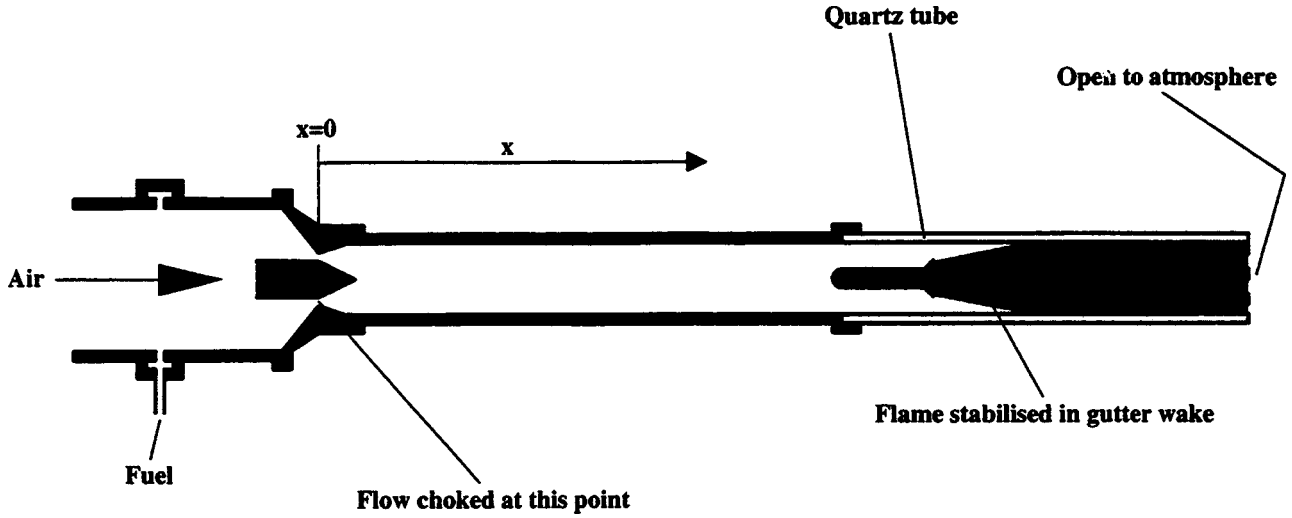


Fig. 1 The experimental rig of Langhorne [8]

The problem is formulated in cylindrical polar coordinates, taking advantage of the axisymmetric geometry to eliminate the azimuthal direction and consider only the axial and radial directions. A Reynolds-averaged approach is taken, using Favre density-weighted ensemble averaging, and a second moment closure model is applied to calculate the Reynolds stresses ([9]). Turbulent scalar transport is modeled using a standard gradient transport hypothesis. The resulting transport equation for the Favre-mean reaction progress variable is

$$\frac{\partial}{\partial t}(\overline{\rho}\widetilde{c})+\frac{\partial}{\partial x_k}(\overline{\rho}\widetilde{u}_k\widetilde{c})=\overline{w}+\frac{\partial}{\partial x_k}\left(\frac{C_\mu}{Sc}\overline{\rho}\frac{\widetilde{k}^2}{\widetilde{\epsilon}}\frac{\partial\widetilde{c}}{\partial x_k}\right), \tag{2}$$

where $\overline{w}$ is the mean turbulent reaction rate, expressed as the production rate of reaction progress variable per unit volume. This quantity is modeled using a laminar flamelet concept, in which the turbulent flame is assumed to consist of an ensemble of thin, highly wrinkled surfaces that separate unburned reactants from fully burned products. The surface has the local structure of a laminar flame, albeit strained and curved by the local turbulent flow. The flamelet model leads to the simple expression

$$\overline{w}=\rho_R u_l^0 I_0 \Sigma, \tag{3}$$

where the group $\rho_R u_l^0 I_0$ represents the mean reaction rate per unit surface area and Σ is the flame surface area per unit volume. The reactant density ρ_R is a known quantity, while the planar unstrained laminar burning velocity u_l^0 is a thermochemically well-defined quantity obtained by means of an empirical correlation ([10]). The quantity I_0 accounts for the change in the laminar burning velocity due to turbulent straining and is given by

$$I_0=\int \frac{u_l(a)}{u_l^0(0)}P(a)da, \tag{4}$$

where the strained laminar flame speed $u_l(a)$ is evaluated from laminar flame computations with detailed chemistry ([11]) and the probability density function $P(a)$ has been obtained from direct numerical simulation data ([12]). Finally, the flame surface density Σ is modeled using a semi-empirical correlation calibrated against the stagnation-plate flame data of Cheng and Shepherd [13].

The model is implemented in the CFD code TARTAN ([14]) developed specifically for unsteady combustion applications. The equations are discretized in space using the curvature compensated convective transport (CCCT) scheme of Gaskell and Lau [15] which is third order accurate in smooth regions of the solution and is flux limited to second order near steep gradients in order to guarantee boundedness. No other numerical smoothing is required. Euler implicit time differencing is employed, and time advancement is by means of the PISO algorithm originally proposed by Issa [16], and subsequently modified to deal with the large density changes commonly encountered in combustion problems. The algorithm begins with an implicit predictor step which updates all quantities except the pressure. There follows a corrector step, in which the pressure Poisson equation is solved implicitly while all other quantities are treated explicitly. A second corrector step repeats the process in order to achieve a second order operator splitting error. The stability of the algorithm is governed mainly by the convective Cousant-Friedrichs-Lewy (CFL) criterion and the time step is automatically adjusted to 80 percent of the CFL-defined value.

Results and Discussion

The test case is the experiment of Langhorne [8] (see Fig. 1) for which a great deal of data exists covering both forced and self-exited oscillations. The apparatus consists of a long straight cylindrical tube with inside diameter 70 mm, supplied with premixed fuel and air through a choked aperture at the entrance. An axisymmetric bluff-body flameholder is located 0.74 m from the entrance

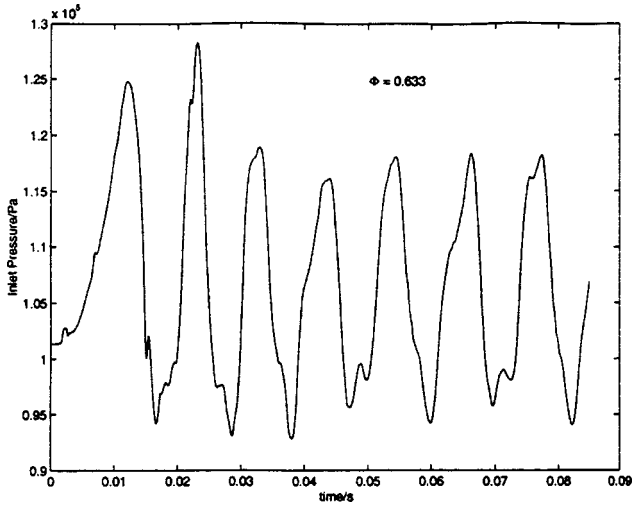


Fig. 2 Pressure at the flameholder versus time, equivalence ratio=0.633

and the overall length of the tube in the case considered here is 1.48 m. The exit from the tube is open to the atmosphere. The geometry is axisymmetric, and the flameholder consists of a short cylindrical body having a rounded nose and a flared lip at the tail providing 25 percent blockage. The fuel is ethylene, and is supplied at fuel-air mass equivalence ratios in the range 0.63 to 0.70. The inlet temperature was 288 K and the mean velocity of the fuel-air mixture upstream of the flameholder was 31.3 m/s. The experimentally observed flame is axisymmetric and exhibits no tendency toward rotational modes of oscillation.

A two-dimensional computational grid was generated for the upper half of the full length of the tube, having a total of 36×230 cells including 70 cells upstream of the flameholder. The grid was spaced non-uniformly in both axial and radial directions, with a high concentration of cells in the region of the flameholder. Tests with half as many cells in each direction revealed no major change in the solution, though the extra grid resolution near the flameholder was found to be important at low equivalence ratio. Tests with reduced time steps were found to produce no changes in the unsteady solution.

Constant values were set for mass flow rate and specific entropy at the entrance, and zero-gradient conditions were set for all variables at the exit except for static pressure which was held constant. In acoustic terms this approximates to choked flow at the inlet and perfect reflection at the outlet. No-slip conditions were set at all walls, using wall functions for turbulence quantities. Convective heat transfer at the walls and radiative heat transfer throughout were neglected. Initial conditions were set for a non-reacting flow, and a solution was obtained in order to establish the main features of the flow field such as boundary layers and recirculation zones. The non-reacting flow field was observed to be steady once converged. Ignition was simulated by artificially increasing the reaction progress variable in a few cells at the lip of the flameholder until a turbulent flame was firmly established. Ignition was then turned off and the calculation was run until any starting transients were eliminated and a limit cycle was reached. A limit cycle was established for all reacting cases without the need for special measures. Each calculation takes approximately three days of CPU time on a Silicon Graphics Octane workstation with a single R10000 processor.

A series of test cases was run for increasing equivalence ratio starting at a value of 0.633. Results for this case are shown in Fig. 2 for the pressure at the lip of the flameholder and Fig. 3 for the velocity at the exit from the tube. Following a short initial tran-

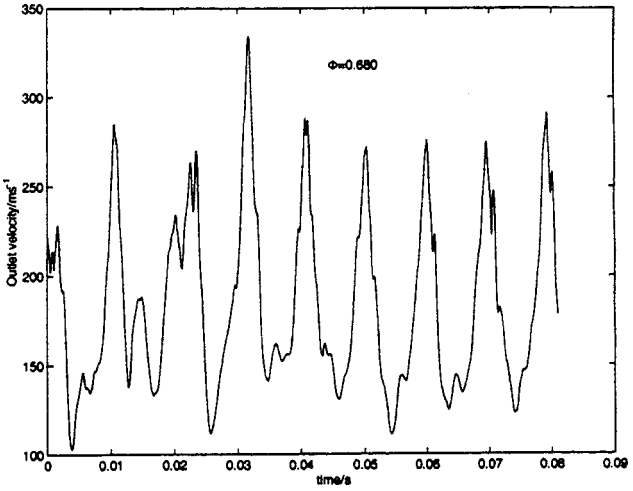


Fig. 5 Velocity at the duct exit versus time, equivalence ratio =0.680

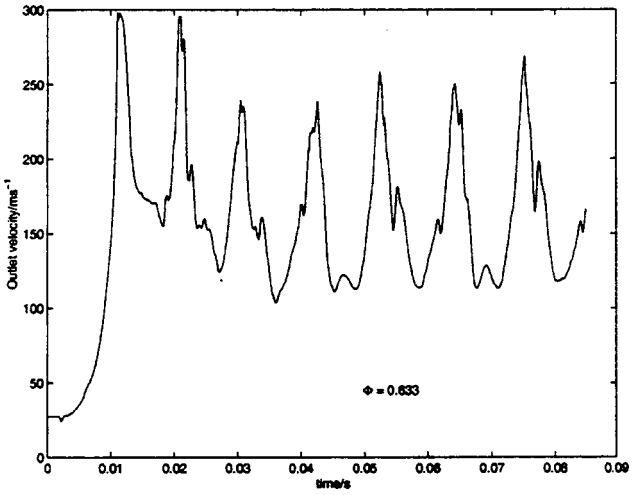


Fig. 3 Velocity at the duct exit versus time, equivalence ratio =0.633

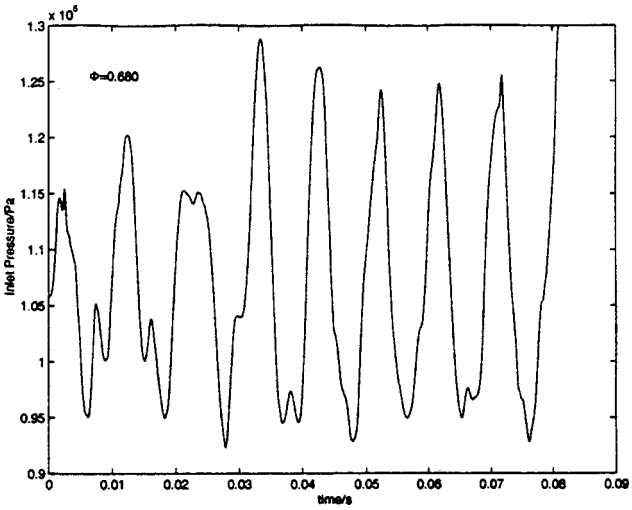


Fig. 4 Pressure at the flameholder versus time, equivalence ratio=0.680

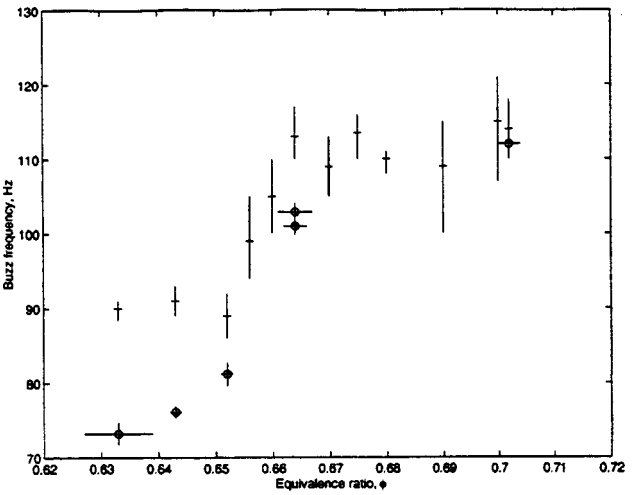


Fig. 6 Frequency of oscillation versus equivalence ratio, showing the transition to self-excited behavior. Circled points denote experimental results.

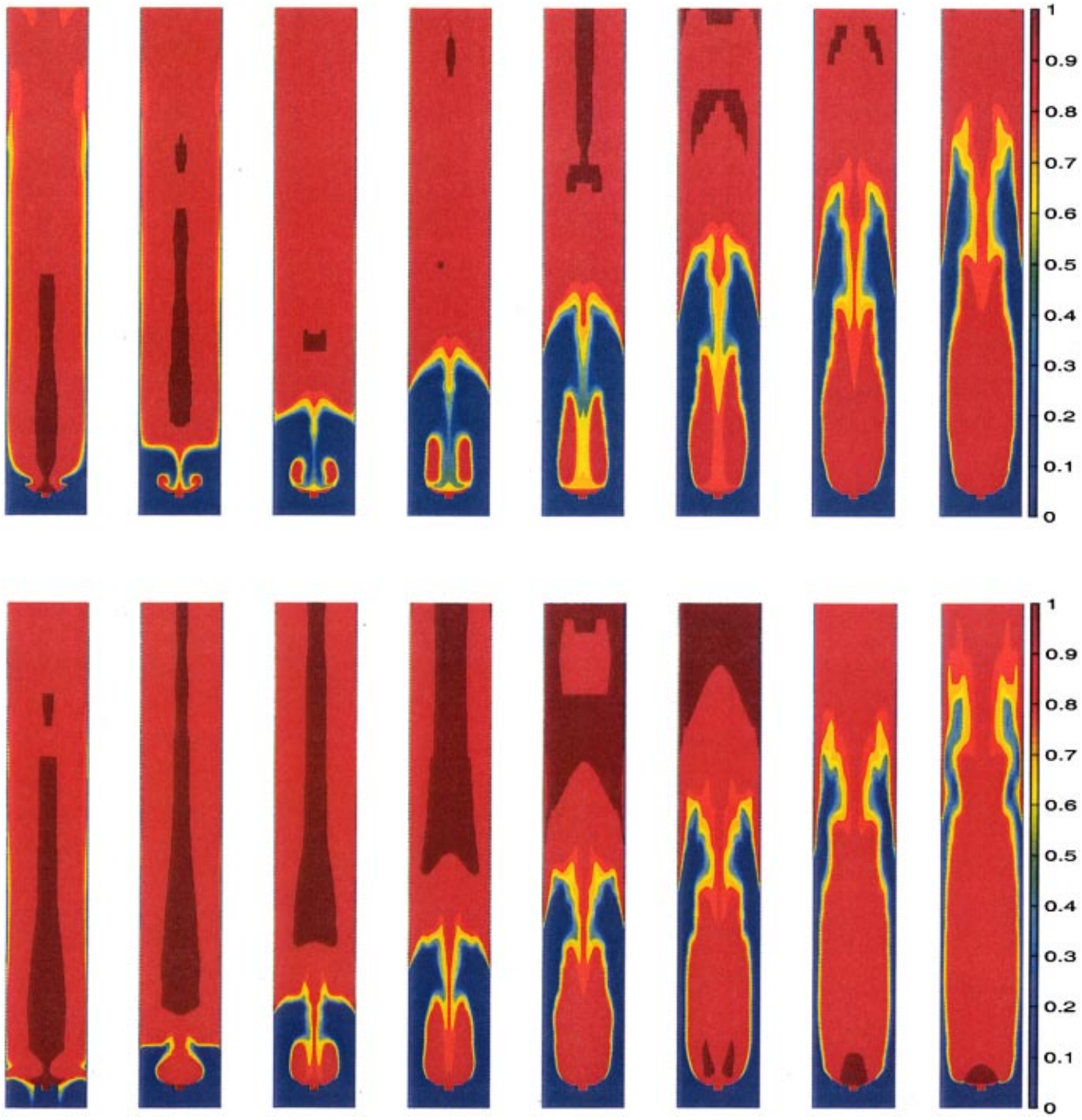


Fig. 7 (**a**) **Contours of reaction progress variable at equal intervals through the cycle, at equivalence ratio=0.633** (**weak oscillation**). (**b**) **Contours of reaction progress variable at equal intervals through the cycle, at equivalence ratio=0.680** (**strong oscillation**).

sient the solution settles down to a stable low-amplitude limit cycle. The frequency of oscillation is about 90 Hz and the magnitude of the pressure fluctuation is about 20 percent. The detailed behavior of the flame through one period of the oscillation is shown in Fig. 7*a*, where contours of reaction progress variable are displayed. Note that the whole tube downstream of the flameholder is shown and that the radial dimension of the tube has been exaggerated for clarity. The flow is from the bottom of the figure and time increases from left to right in intervals of 1 ms. Only the domain between the leading edge of the flameholder and the exit from the tube is shown, and the solution has been reflected in the centerline in order to emphasize the symmetry. The sequence shows the flame immediately behind the flameholder becoming distorted by a vortex shed from the lip. A major portion of the flame is detached from the flameholder and is convected downstream, leaving behind a small kernel of hot burning gas. The flame is able to re-establish itself in the growing recirculation zone, burning vigorously through the reactant mixture trapped in the vortex, and propagating outward into the surrounding high-velocity gas. The burned products are swept downstream as the flame consumes the remaining unburnt material in the tube and the sequence is ready to begin again.

By an equivalence ratio of 0.688 the system has undergone a transition. The frequency deduced from the inlet pressure trace,

shown in Fig. 4, and outlet velocity trace, shown in Fig. 5, has risen to about 110 Hz, while the magnitude of the pressure fluctuation is over 30 percent and appears to be increasing. Figure 7*b* shows the detailed behavior of the flame for this case, displayed as before but with an interval of 0.8 ms. The sequence of events is broadly similar to that found at lower equivalence ratio, although the re-establishment of the flame following the initial vortex interaction is somewhat faster here. The key difference in flame behavior between the two cases is the propagation of the flame upstream of the flameholder in the higher-amplitude case. This may be seen in the final image of the sequence in Fig. 7*b*, while the first image shows the situation as the flame is in the process of being swept back downstream. Propagation of the flame upstream of the flameholder has been found experimentally to be an excellent indicator of the higher-amplitude self-excited behavior, and is also in agreement with theoretical findings.

The ability of the CFD calculation to capture the transition between the two different types of self-excited oscillation is illustrated in Fig. 6, which shows the computed and measured frequency plotted against equivalence ratio. The experimental data points are indicated by circles, the computed points by crosses, and error bars are shown corresponding to the estimated error in the equivalence ratio in the measurements, and to the estimated error in the frequency for both. In the experimental data there is a clear transition from low frequency to high frequency at an equivalence ratio of about 0.66, and this is matched very well by the computations. In general there is very good agreement between computed and measured frequency at higher equivalence ratios, but there is some discrepancy at equivalence ratios below the transition point, i.e., within the weak or lower-amplitude regime, where the computed results produce higher frequencies than were measured. The computed amplitude of the oscillation is in reasonable agreement with experiment for the higher equivalence ratios, but again the agreement is less satisfactory below the transition point where amplitudes are overpredicted. Both of these discrepancies will be investigated in future work, which will also examine the precise mechanism of the coupling between pressure, velocity, and heat release.

Conclusion

A series of CFD calculations has been carried out in order to assess the capability of the technique for the prediction of self-excited acoustically coupled combustion instabilities. An unsteady Reynolds-averaged approach has been taken, using Reynolds stress turbulence modeling and laminar flamelet combustion modeling. Comparison with experiment has revealed a high level of agreement with the character and the frequency of the observed oscillation. The CFD technique has also been shown to capture the transition point between two distinct types of self-excited oscillation, at a fuel-air equivalence ratio that is in good agreement with the experimental data.

Acknowledgment

The financial support of Hitachi Europe Ltd. is gratefully acknowledged.

Nomenclature

Italic

a = Flame tangential strain rate
c = Reaction progress variable
C_μ = Model constant, value=0.09
I_0 = Strain correction factor
k = Turbulence kinetic energy
$P(a)$ = Probability density function (pdf) of a
Sc = Turbulent Schmidt number, value=1.0
t = Time
u_i = Velocity vector component
u_l = Laminar burning velocity
x_i = Position vector component
w = Reaction rate
Y = Mass fraction

Greek

ϵ = Turbulence energy dissipation rate
ρ = Density
Σ = Flame surface area per unit volume

Subscripts

F = Fuel
P = Products
R = Reactants

Diacritics

$\widetilde{c}$ = Favre average of c
$\overline{c}$ = Reynolds average of c

References

[1] Smith, C. E., and Leonard, A. D., 1997, ''CFD Modelling of Combustion Instability in Premixed Axisymmetric Combustors,'' ASME Paper No. 97-GT-305.
[2] Steele, R. C., Cowell, L. H., Cannon, S. M., and Smith, C. E., 1999, ''Passive Control of Combustion Instability in Lean Premixed Combustors,'' ASME Paper No. 99-GT-052.
[3] Brookes, S. J., Cant, R. S., and Dowling, A. P., ''Modelling Combustion Instabilities Using Computational Fluid Dynamics,'' ASME Paper No. 99-GT-112.
[4] Menon, S., and Jou, W-H., 1991, ''Large-Eddy Simulation of Combustion Instability in and Axisymmetric Ramjet Combustor,'' Combust. Sci. Technol., **75**, pp. 53–72.
[5] Möller, S. I., Lundgren, E., and Fureby, C., 1996, ''Large Eddy Simulation of Unsteady Combustion,'' 26th Symposium (International) on Combustion, The Combustion Institute, pp. 241–248.
[6] Murota, T., and Ohtsuka, M., 1999, ''Large-Eddy Simulation of Combustion Oscillation in the Premixed Combustor,'' ASME Paper No. 99-GT-274.
[7] Dowling, A. P., 1999, ''A Kinematic Model of a Ducted Flame,'' J. Fluid Mech., **394**, pp. 51–72.
[8] Langhorne, P. J., 1988, ''Reheat Buzz: An Acoustically Coupled Combustion Instability. Part I: Experiment,'' J. Fluid Mech., **193**, pp. 417–443.
[9] Cant, R. S., and Bray, K. N. C., 1989, ''A Theoretical Model of Premixed Turbulent Combustion in Closed Vessels,'' Combust. Flame, **76**, pp. 243–263.
[10] Abu-Orf, G. M., and Cant, R. S., 1996, ''Reaction Rate Modelling for Premixed Turbulent Methane-Air Flames,'' Proc. Joint Meeting of the Portuguese, British, Spanish and Swedish Sections of the Combustion Institute, Madeira.
[11] Cant, R. S., Bray, K. N. C., Kostiuk, L. W., and Rogg, B., 1994, ''Flow Divergence Effects in Strained Laminar Flamelets for Premixed Turbulent Combustion,'' Combust. Sci. Technol., **95**, pp. 261–276.
[12] Bray, K. N. C., and Cant, R. S., 1991, ''Some Applications of Kolmogorov's Turbulence Research in the Field of Combustion,'' Proc. Roy. Soc. Lond., **A434**, pp. 217–240.
[13] Cheng, R. K., and Shepherd, I. G., 1991, ''The Influence of Burner Geometry on Premixed Turbulent Flame Propagation,'' Combust. Flame, **85**, pp. 7–26.
[14] Watkins, A. P., Li, S-P., and Cant, R. S., 1996, ''Premixed Combustion Modelling for Spark Ignition Engine Applications,'' Society of Automotive Engineers, SAE Paper No. 961190.
[15] Gaskell, P. H., and Lau, A. K. C., 1988, ''Curvature-Compensated Convective Transport: SMART. A New Boundedness-Preserving Transport Algorithm,'' Int. J. Numer. Methods Fluids, **8**, pp. 617–641.
[16] Issa, R. I., 1986, ''Solution of the Implicitly Discretised Fluid Flow Equations by Operator Splitting,'' J. Comput. Phys., **62**, pp. 40–65.

Coupled and Uncoupled CFD Prediction of the Characteristics of Jets From Combustor Air Admission Ports

J. J. McGuirk

A. Spencer

Department of Aeronautical and Automotive Engineering,
Loughborough University,
Loughborough, Leicestershire, LE11 3TU, UK
A.Spencer@lboro.ac.uk

The paper focusses attention on alternative approaches for treating the coupling between the flow in the annulus supply ducts and the jets which enter combustor primary and dilution zones through air admission ports. Traditionally CFD predictions of combustor flows have modeled this in a very weakly coupled manner, with the port flow conditions being derived from 1D empirical correlations and used as boundary conditions for an internal-flow-only combustor CFD prediction. Recent work by the authors and others has introduced the viewpoint that fully coupled external-annulus–internal-combustor predictions is the way forward. Experimental data is gathered in the present work to quantify the strength of the interaction between annulus and core flows, which ultimately determines the jet characteristics at port exit. These data are then used to illustrate the improvement in the prediction of port exit jet characteristics which is obtained by adopting fully coupled calculations compared to the internal-flow-only approach. As a final demonstration of the importance of a fully coupled approach, isothermal calculations are presented for a single sector generic annular combustor. These show that quite different primary zone flow patterns are obtained from the two approaches, leading to considerable differences in the overall mixing pattern at combustor exit.
[DOI: 10.1115/1.1362319]

Introduction

It is rare that combustor CFD calculations include (simultaneously) both external as well as internal aerodynamics. The vast majority of computational predictions of combustors available in the open literature (and in all probability those carried out in industry) have generally been performed in what might be described as an uncoupled (or at best a one-way coupled) two-step mode. The external flow from compressor exit to the various entry routes into the combustor is considered first, typically using a quasi one-dimensional (1D) network analysis based on empirical correlations of liner hole discharge coefficients, e.g., Adkins and Gueroui [1]. Karki et al. [2] have presented a slightly improved version of this where the external flow was calculated using a 2D CFD method. This first step provides (at most) information on the mass flow splits and jet entry angles at the various ports in the combustor liner surface. These then act as the boundary conditions (hence one-way coupled) for the second step, which comprises the complete internal combustor flowfield prediction.

McGuirk and Spencer [3,4], have pointed out that this practice may be deficient in at least two important aspects. No information is provided on velocity profile shapes, or turbulence conditions, at the port entry location. It is likely that, under some conditions at least, the characteristics of the jets emerging from air entry ports are sensitive to profile and turbulence information. Since the internal combustor flow pattern is known to be highly sensitive to jet trajectory, penetration, etc., then the above weakly coupled practice is not to be recommended.

An early example of a combustor calculation, which contained full coupling of both internal and external flows, was provided by Manners [5]. However, the limitations of the orthogonal grid used meant that the port shapes were not properly resolved. Coupled calculations for a geometrically simple annulus/internal flow which removed this defect were provided by McGuirk and Spencer for plain round holes ([3]), and chuted ports ([4]), to illustrate the sensitivity of predicted jet behavior to coupled CFD treatment. More recently, Bain et al. [6] have examined flow coupling effects on the jet-in-crossflow flowfields found in rich burn/quick mix/lean burn combustors. Their conclusions support the position adopted above, i.e., that the strong coupling between jet and mainstream flow cannot be captured by the commonly used specified jet entry boundary conditions. Crocker et al. [7] have also arrived at the same conclusion in their predictions of a model gas-turbine combustor, which encompassed the whole flow from compressor exit to turbine inlet. However, no quantitative comparison was given between uncoupled and fully coupled solutions, and no experimental data were available for the model combustor used to confirm the adequacy of the predictions.

Although belief in the importance of fully coupled combustor modelling is evidently growing, detailed experimental confirmation of the benefits of coupled calculations is still lacking. The present work was carried out with this end in mind, i.e., to gather experimental data on jet characteristics in a generic external/internal flowfield. Experimental work to generate such data in a specially designed test facility is described below. This is followed by a CFD study to illustrate the differences obtained, for the experimental setup, between boundary condition-driven and fully coupled predictions. Finally, the CFD study is extended to indicate the effect on predicted isothermal internal flow patterns in a typical annular combustor of a switch from decoupled to fully coupled treatment.

Experimental Methods

An experimental facility containing a generic annulus flow/port flow/internal flow combination has been constructed to allow detailed surveys to be made of discharge coefficients, jet exit veloc-

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Munich, Germany, May 8–11, 2000; Paper 00-GT-125. Manuscript received by IGTI February 2000; final revision received by ASME Headquarters January 2001. Associate Editor: M. Magnolet.

ity conditions and associated jet/internal flow interactions for a range of port shapes and annulus/core flow parameters. A water flow facility was selected to avoid seeding problems for the LDA measurements to be undertaken. Reynolds numbers of all flows were sufficiently large ($>2\times10^4$) to provide proper simulation of airflow at typical combustor operating conditions. Figure 1 presents a schematic of the test section of the isothermal, vertically flowing, constant-head waterflow rig. The test section comprises a pair of circular Plexiglas tubes of 600 mm length; the tubes are held concentric using a set of NACA 0015 struts at the upper and lower ends. The flow is supplied to the outer and inner tubes from a constant head tank to provide annulus and internal (core) flows. The core flow pipe contains a single row of six 20 mm diameter ports approximately halfway along its length. Changing the inner pipe allows port geometry to be varied; sharp-edged, chuted, and stepped round holes, and D-shaped holes have been investigated (Spencer [8]). Valves in both upstream (supply) and downstream (bleed) parts of the annulus flow circuit are used to control the percentage of the annulus flow which passes through these holes, to form a row of radially inflowing jets interacting with the core flow. This allows conditions typical of primary zone jets (a jet/core flow velocity ratio of 5) and dilution zone jets (velocity ratio of 2) to be established. A square acrylic water jacket was arranged to surround the entire working section, as recommended by Bicen [9] to reduce refraction effects at the air-acrylic interface.

The coordinate system and nomenclature adopted in the experiment are given in Fig. 2.

A simple 1D, low-power (10 mW) LDA system was available to gather mean velocity and turbulence data. The system could be used in either forward- or backscatter mode. The former was adequate for the majority of measurements. Occasionally backscatter was used since this gave the advantage of common receiving and transmitting optics. Hence, when traversing through regions of substantial refraction, constant realignment was not required. Data

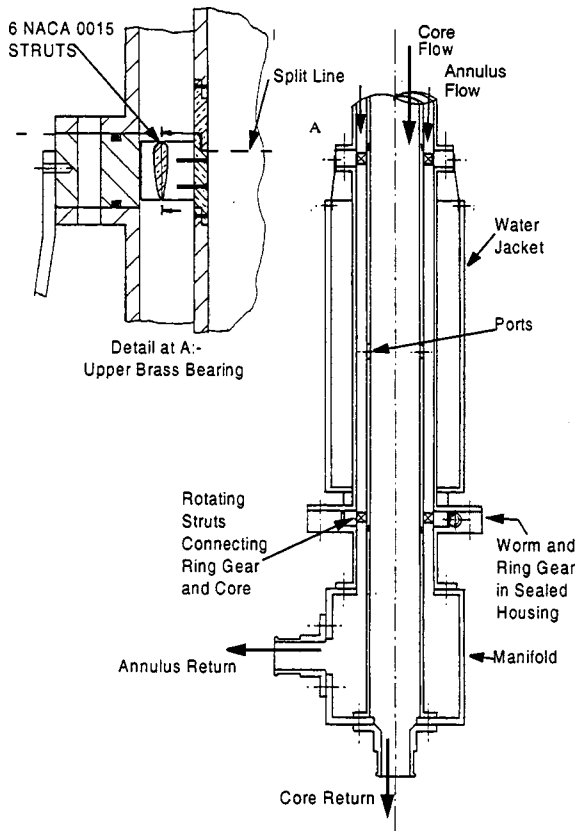


Fig. 1 Schematic of the test section

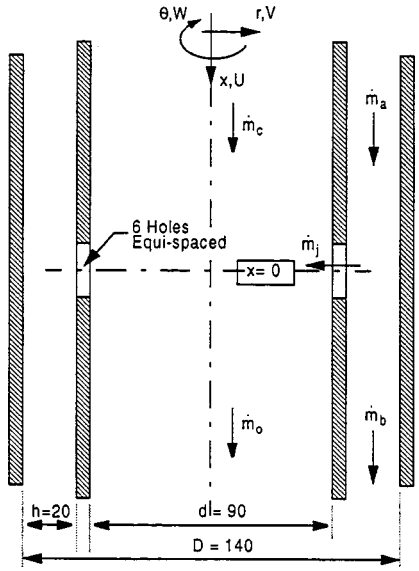


Fig. 2 Coordinate system for the test section (dimensions in mm)

rates in forward- and backscatter modes were typically 20 and 2 kHz, respectively. The entire inner pipe was rotated about its axis using a worm and ring gear arrangement located at the lower strut position. This obviated the need to rotate the laser optical axis about the tube axis to measure velocities on different azimuthal planes. The entire LDA system was mounted on a milling table traverse to give three axes of translation. These arrangements allowed all three mean velocity components and associated turbulent stresses to be measured in all planes of interest with the LDA control volume located to an accuracy of ±0.01 mm. More detail of the instrumentation used may be found in Spencer [8].

The velocity and turbulence measurements presented below are for plain round holes and were taken on the following planes.

1 Inlet conditions (for use in CFD calculations) were measured on plane $x=-150$ mm and confirmed to be axisymmetric to within one percent,
2 $\theta=0°$ plane (through a port diameter) at eight axial locations ($x=-75$, -50, -30, -10, -5, 5, 15, and 40 mm), and around 40 radial points across whole diameter of test section,
3 $\theta=0$ deg plane and at port exit ($r=45$ mm) for $-10<x<10$ mm.

Numerical Methods

The computational approach adopted in the present work was identical to that used in our earlier studies ([3,4]), where full details are given, so only a summary description is provided here (see also Spencer [8]). The code used was based on 3D, cell-centered, finite-volume, implicit solution of the Reynolds-averaged Navier-Stokes equations using the standard high Reynolds number two equation k-ε model (Launder and Spalding [10]) with conventional wall-functions for the wall treatment. A SIMPLE-based pressure-correction formulation ([11]) was combined with Rhie-and-Chow-based pressure smoothing to avoid pressure uncoupling of the colocated variables (see Ferziger and Peric [12]). In order to treat complex-shaped domains, the Cartesian forms of the transport equations were transformed into a general 3D nonorthogonal curvilinear system. Structured meshes were generated to fit the geometries of interest using the elliptic pde method of Thompson [13]. The method described in Ref. [4] was used to extend these ideas to fitting the shape of port geometries. Details on the particular grids used to obtain calculations reported here are given in the results section; grid independence

Table 1 Main flow parameters for primary and dilution test cases

Flow Condition	Primary	Dilution
$R(V_j/U_c)$	5.0	2.0
$B(\dot{m}_b/\dot{m}_a)$	50 percent	20 percent
Re_j (subscripts as Fig. 2)	$>2.4\times10^4$	$>2.4\times10^4$

tests are reported in Spencer [8]. For comparison with experiments, the initial boundary conditions were taken from the measurements, as detailed under (1) in the last section. For the combustor calculation, boundary conditions are described fully below.

Results

Experimental Data. Table 1 provides a list of the main flow parameters for the two cases studied in detail, which were aimed at providing typical characteristics of a jet in the combustor primary zone (velocity ratio $R=5$, Bleed flow ratio $B=50$ percent) and also the dilution zone ($R=2$, $B=20$ percent). Figures 3 and 4 capture the exit profiles which characterize the jet's entry from the annulus flow through the ports into the cross-flowing core flow. For a primary jet (Fig. 3) the through port radial velocity is quite nonuniform, reaching a peak value in this profile at the rear of the hole some 70 percent greater than the bulk velocity through the port (V_j). At the front of the port the velocity is negative, indicating that separation has taken place and flow is actually directed from the core flow into the hole over a small region. The fact that the port is fed from an annulus with considerable bleed flow and not from a plenum is indicated by the existence of a significant axial component of velocity in the jet—peaking at 60 percent of V_j at the front of the port. These data may be combined to indicate the jet angle at port exit, also shown in Fig. 3 which varies from 60° to 90° over the port (ignoring the negative angles at the leading edge). The turbulence measurements show a low turbulence core (around 5 percent turbulence intensity), but high turbulence is generated at both upstream and downstream edges; peak turbulence intensity is around 40 percent at the leading edge (associated with the small recirculation inside the port), but even at the downstream edge the turbulence intensity rises to 30 percent, as a consequence of the shear occurring in the portion of the annulus flow drawn into the flow through the port. With a reduction in velocity ratio and bleed to dilution zone conditions ($R=2$, $B=20$ percent), the jet characteristics change in several notable ways (Fig. 4). First the upstream edge recirculation is now not present, as evidenced in radial velocity, turbulence energy and jet angle profiles. However, the radial profile shows a more distinct downstream peak than previously; associated with this is evidence of reverse flow in the annulus outside the port, as the axial velocity at downstream port exit is negative. This implies that at low bleed ratios the port cannot capture all of its flow from upstream and some of the flow enters the port from the downstream annulus.

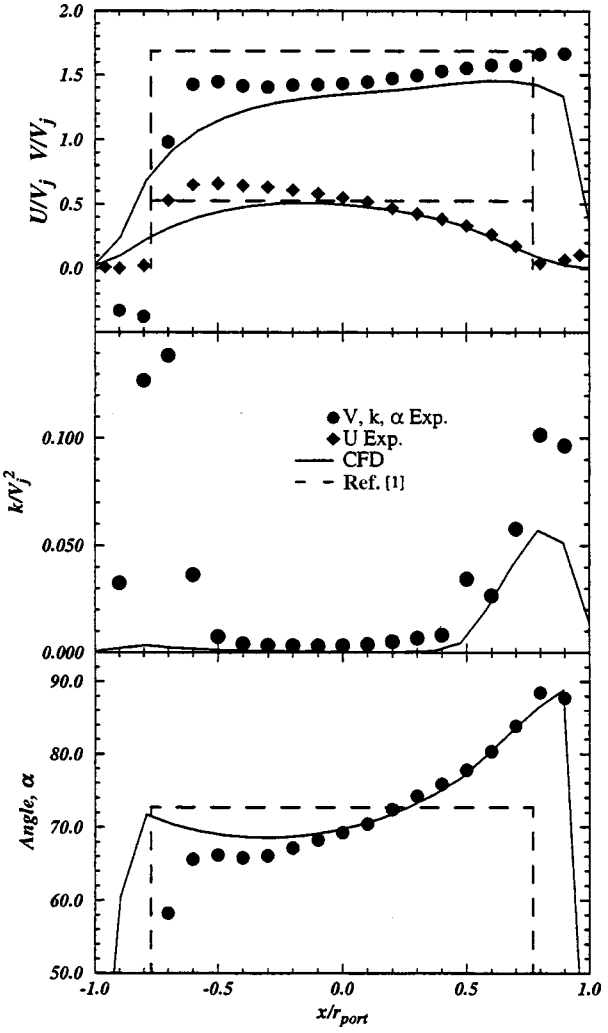


Fig. 3 Primary jet characteristics

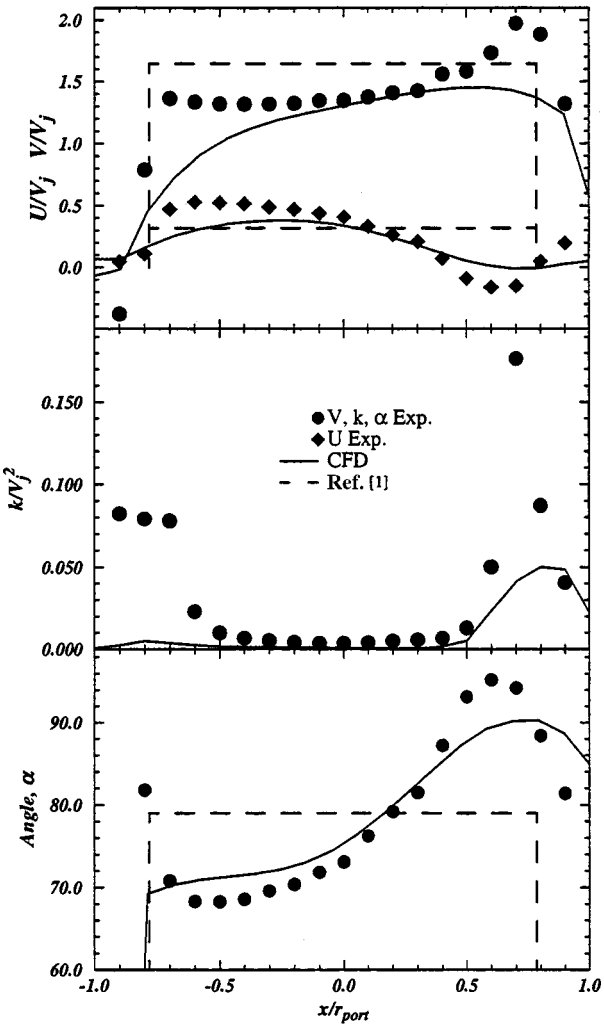


Fig. 4 Dilution jet characteristics

The measurements taken within the core-flow allow the conse-

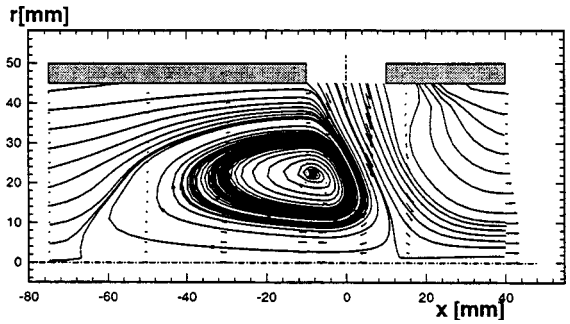


Fig. 5 Measured internal flow-particle tracks. Primary configuration ($R=5$, $B=50$ percent).

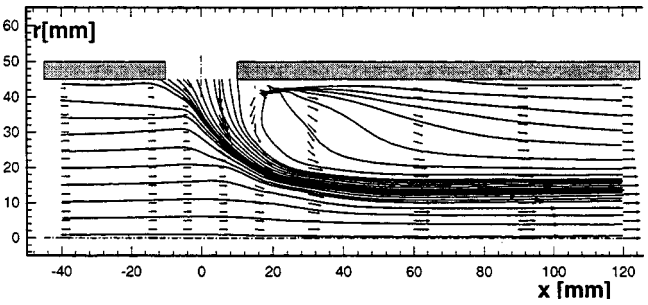


Fig. 6 Measured internal flow-particle tracks. Dilution configuration ($R=2$, $B=20$ percent).

quences of these different jet characteristics to be captured. These are shown in Figs. 5 and 6 for primary and dilution flow conditions. These figures use particle paths tracked through the measured velocity field in the $\theta=0°$ plane to visualize the flow pattern. These plots establish the jet trajectories very clearly and confirm that realistic velocity fields have been set up to simulate combustor primary and dilution zones.

Calculations for Experimental Annulus/Core-flow Conditions. The CFD methodology described earlier has been applied to the prediction of the experiments described above. Fully coupled calculations, including both annulus and core flow in the CFD mesh, have been carried out, as well as decoupled calculations where only the internal core flow field was predicted, the port flow boundary conditions (mean velocity and jet angle) being taken from the empirical calculations of Adkins and Gueroui [1] as described in the Introduction. The values adopted as jet boundary conditions are shown for the primary port configuration in Fig. 3 as dashed lines. It can be seen that the 1D correlations assume that the jet issues through a smaller-than-actual port with an effective diameter ($A_{\text{eff}}=C_d\times A_{\text{port}}$) to take into account the vena-contracta effect. This practice is typical of that used in industry. The grid used for the fully coupled calculations is shown in Fig. 7 (the internal only calculation is obviously a subset of this). Grid refinement studies were carried out (details reported in Spencer [8]) and grid independence was confirmed on the finest grid used of $100\times40\times30$ (axial, radial, azimuthal). Inlet boundary conditions for the annulus and core flow were taken from the measured values at $x=-150$ mm (see Spencer [8]). Figures 8(*a*) and 8(*b*) present velocity vectors on the $\theta=0°$ plane taken from these predictions (for $R=5$) for coupled (a) and internal only (b) approaches. Noticeable differences result in terms of jet spreading and the upstream size, shape and strength of the recirculation zone. For example, the upstream recirculation of the internal only prediction is greater than for the coupled prediction. Although this might be seen as moving the prediction closer to the experimental results, it will be shown below that this is because the 1D empirical correlation substantially overestimates the radial inflow velocity over the whole of the upstream half of the port geometry; a longer recirculation is inevitable but for quite the wrong reason.

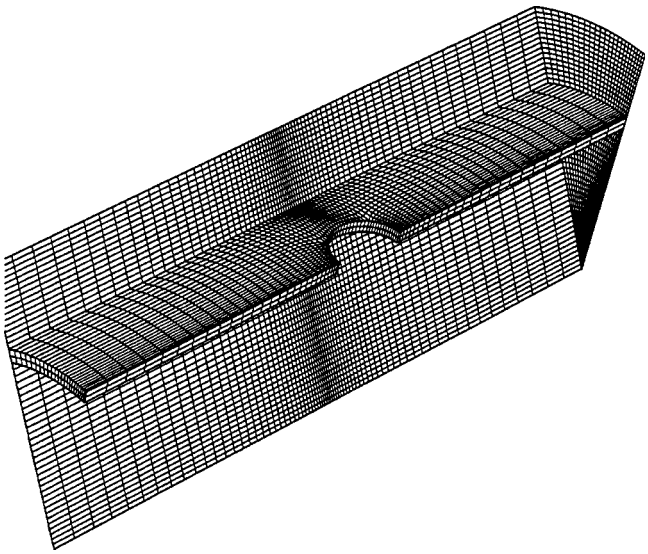


Fig. 7 Computational mesh of rig geometry

Figure 8 shows that the inclusion of the annulus flow does exert an influence on internal core-flow simulation. The coupled calculation indicates clearly the strong shear in the annulus which leads to turbulence generation at the port downstream edge; indeed, for these conditions there is a noticeable region of reversed flow predicted in the annulus just above the downstream port edge.

If the predicted jet exit profiles from the coupled calculation and the 1D correlation deduced conditions are analyzed for the $\theta=0°$ plane, then the results are as indicated in Figs. 3 and 4 via solid (coupled CFD) lines (dashed lines indicate the values predicted by the 1D correlation of Adkins and Gueroui [1]). In general the coupled calculation agrees much better with the LDA data

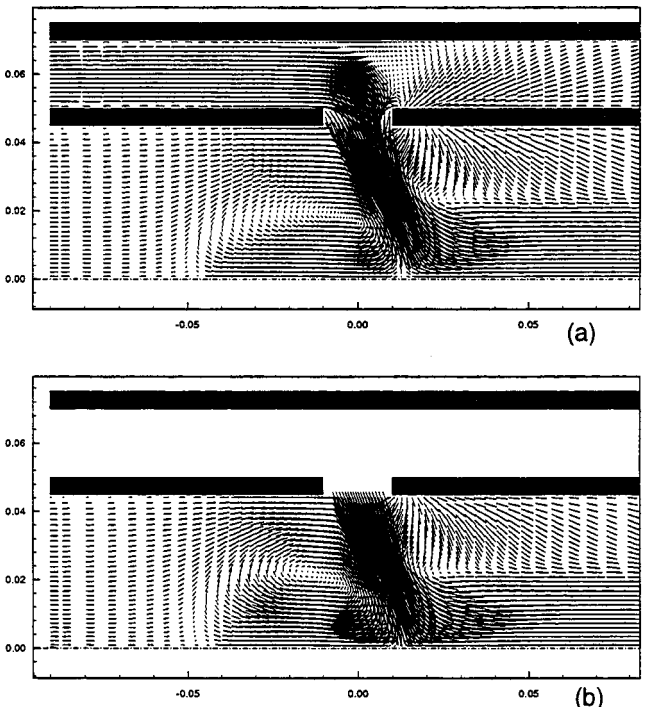


Fig. 8 Predicted velocity field. (*a*) Coupled and (*b*) internal only.

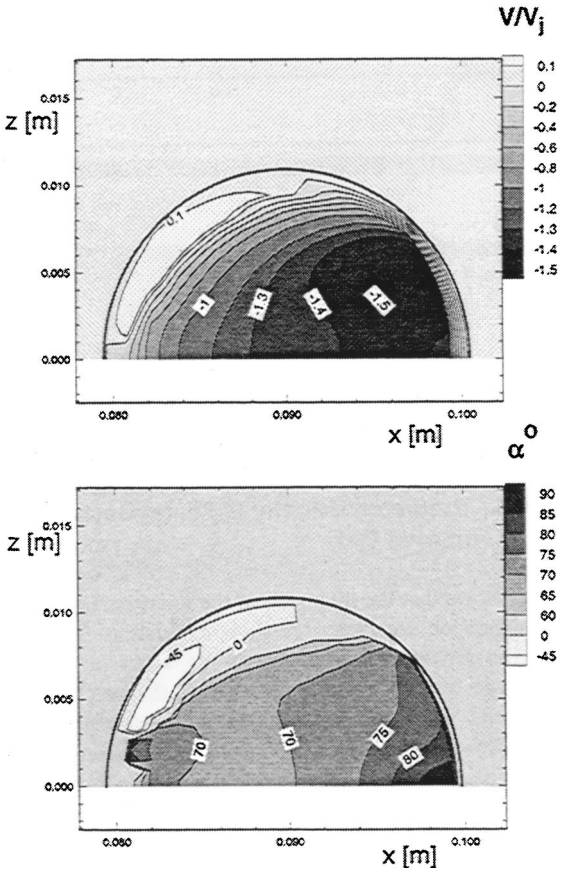


Fig. 9 Primary jet through port velocity and flow angle-coupled CFD prediction

than the 1D correlation. The turbulence energy profile is underpredicted in terms of the intensity, particularly at the upstream edge. (However, it must be noted that the 1D correlations give no information on jet turbulence levels and a sensible guess must be made to complete the boundary conditions for the internal only calculation.) For the primary jet case this is clearly linked to the lack of any recirculation in the predictions. This is not, however, as bad as it seems. Examination of Fig. 9 (which shows coupled CFD predictions over port exit for radial velocity and jet angle for primary jet conditions) reveals that the CFD prediction has resulted in a core-to-annulus direction flow within the port, although this has not (quite) reached the $\theta=0°$ plane. It is suspected that this is a local grid resolution problem which is worsened across the $\theta=0°$ diameter where the grid cells are at their largest due to the use of a structured grid to represent the circular port. The strength of this reverse flow is quite marked (10 percent of V_j) and in general the plots in Fig. 9 underline how crude a single value of radial velocity or jet angle is for this important boundary condition which drives the internal combustor flow.

Calculations for Internal/External Flow in Generic Annular Combustor Geometry. As a final demonstration of the significance of fully coupled versus internal-only CFD predictions, a generic external/internal aerodynamic calculation of an annular combustor is presented. The situation modelled corresponds to isothermal flow in a 15° single burner sector of an annular combustor (i.e., 24 burners). Experimental data were available for the velocity profiles entering the inner and outer annuli of this combustor, so the feed annuli from the combustor heatshield onwards were modelled in the grid. Similarly, the air-swirler characteristics were also available and hence these could be placed as boundary conditions on appropriate regions of the heatshield. The

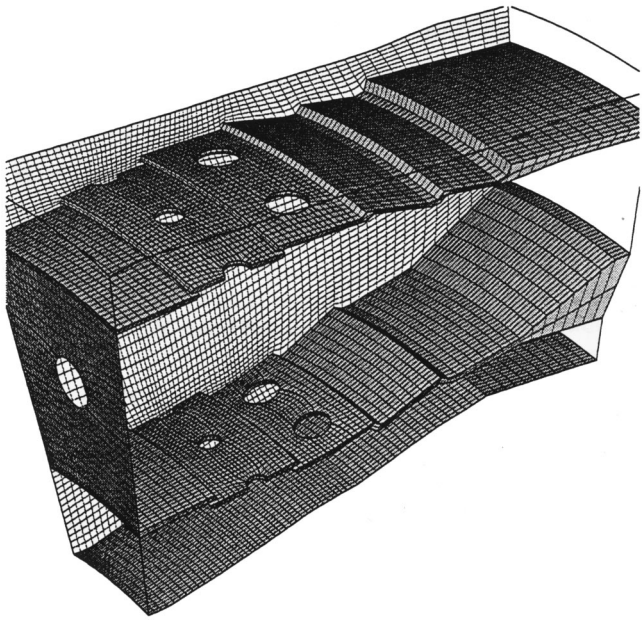


Fig. 10 Modelled combustor sector for internal/external aerodynamic study

cooling rings were captured by the grid used and boundary conditions for the flows entering these features were also included. The flow splits for inboard and outboard primary and dilution jets were *not* fixed in the fully coupled calculations, but were treated via 1D correlations in the internal only calculation. An indication of the geometry and the grid used is given in Fig. 10; (73×57×36) cells were used in the coupled calculation. This represents the level of mesh resolution likely to be affordable in realistic combustor calculations.

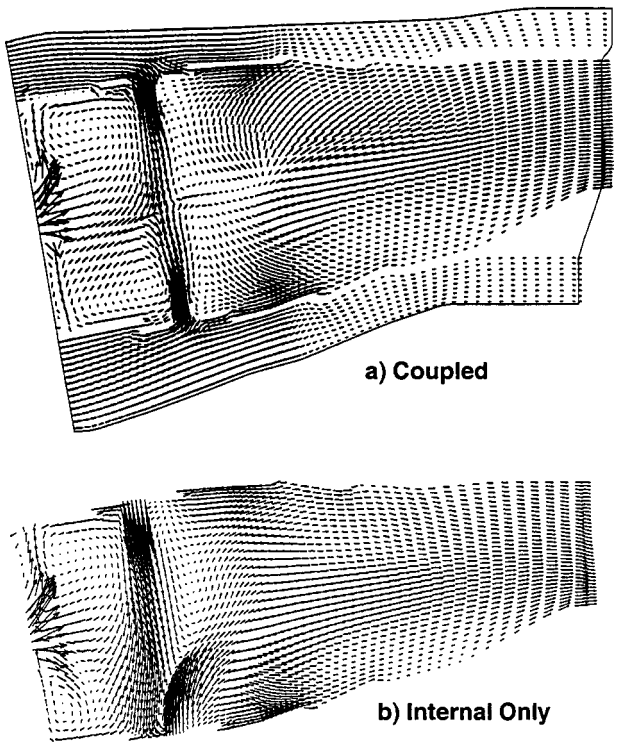


Fig. 11 Predicted velocity fields-primary jets

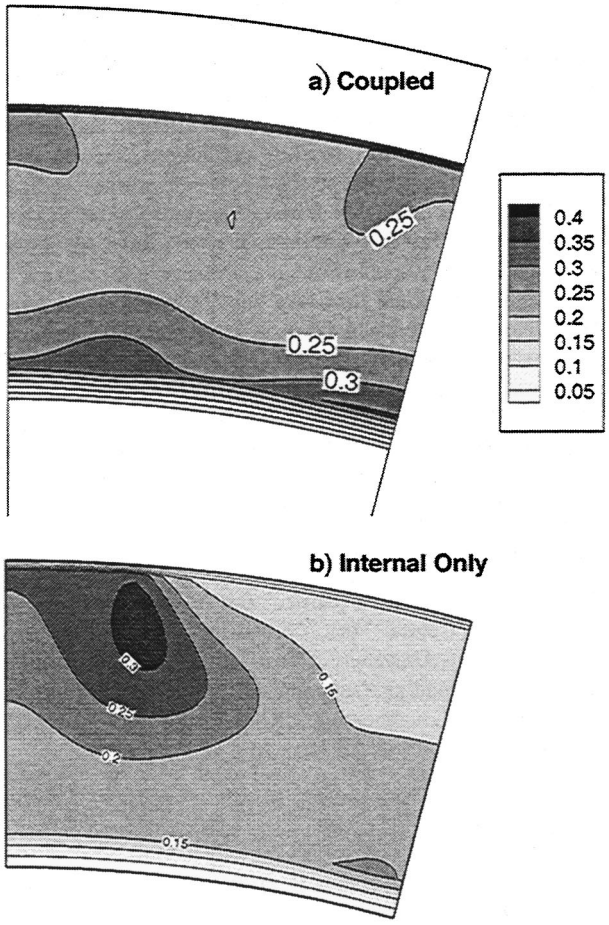


Fig. 12 Concentration on combustor exit plane

Figure 11 shows velocity vector plots on a selected plane, comprising fully-coupled and internal only calculations. Figure 11 displays an axial plane passing through the primary port centreline. Large differences in primary jet penetration can be observed, leading to different impingement behavior and hence a different interaction with the swirler flow and a different primary zone pattern.

To illustrate the consequences of this for mixing, a conserved scalar was released in the swirler flow (concentration = 1.0, zero concentration in all other flows). Plots of this scalar at combustor exit are shown in Fig. 12 (this is effectively the exit, nondimensional temperature traverse in this constant density calculation). The overall magnitude and location of the ''hot spots'' is quite different in the two calculations. For example, in the fully coupled prediction, not only is the overall mixing more uniform, but maximum values are found inboard. For the internal only prediction, the distribution is much peakier—with the maximum value now outboard. Since the predictions described above have shown that the jet exit conditions are modelled better by a fully coupled calculation than an internal only calculation, then only the mixing pattern results in Fig. 12(*a*) are reliable.

Conclusion

Experimental and computational work has been carried out to assess the importance of retaining a strong coupling between external annulus aerodynamics and internal combustor flowfields in CFD predictions of gas-turbine combustors. An isothermal flow rig was designed to create generic annulus and core flow circuits, connected via a single row of ports to create radially inflowing jets. Measurements were taken of the jets that emerged from the ports using an LDA system; two flow configurations were studied, corresponding to primary zone and dilution zone jets. The data indicated a strong effect of the annulus flow details on the profiles of velocity, turbulence level, and jet angle at the port exit, which then determined the jet behavior in the internal core flow. Variations of all these quantities were very nonuniform and could not be adequately represented by simple 1D empirically derived values.

CFD predictions were carried out for the experimental setup using a standard 3D finite-volume method based on structured grids, a pressure-correction method and a high Reynolds number k-ϵ turbulence model. Comparison of fully coupled and internal only predictions showed notable differences in terms of predicted size and strength of jet penetration and upstream recirculation. The fully coupled predictions showed overall good agreement with the detailed port exit profiles, indicating that these predictions were a true reflection of the CFD model performance rather than the internal only results. To illustrate the impact of fully coupled versus internal only predictions for a combustor, isothermal flow in a generic single sector annular combustor was calculated. Once again substantial differences in jet penetration, primary zone flow patterns and mixing levels in the combustor were observed between the two approaches. On the evidence of the work reported here it is recommended that fully coupled calculations of internal/external flow should, where possible, always be carried out in combustor CFD predictions.

Acknowledgments

This work has been carried out within the University Technology Centre in Combustor Aerodynamics at Loughborough University. The authors would like to acknowledge funding, support and useful discussions with colleagues at Loughborough, Rolls-Royce plc., and DERA (Pyestock).

References

[1] Adkins, R. C. and Gueroui, D., 1986, ''An Improved Method For Accurate Prediction Of Mass Flows Through Combustor Liner Holes,'' ASME Paper 86-GT-149.
[2] Karki, K. C., Oechsle, V. L., and Mongia, H. C., 1990, ''A Computational Procedure For Diffuser-Combustor Flow Interaction Analysis,'' ASME Paper 90-GT-35.
[3] McGuirk, J. J., and Spencer, A., 1993, ''CFD Modeling Of Annulus/Port Flows,'' ASME Paper 93-GT-185.
[4] McGuirk, J. J., and Spencer, A., 1995, ''Computational Methods For Modelling Port Flows In Gas-Turbine Combustors,'' ASME Paper 95-GT-414.
[5] Manners, A. P., 1988, ''The Calculation Of The Flows In Gas Turbine combustion Systems,'' Ph.D. thesis, University of London.
[6] Bain, D. B., Smith, C. E., Liscinsky, D. S., and Holderman, J. D., 1999, ''Flow Coupling Effects in Jet-in-Crossflow Flowfields,'' J. Propul. Power, **15**, pp. 10–16.
[7] Crocker, D. S., Nicklaus, D., and Smith, C. E., 1999, ''CFD Modeling of a Gas Turbine Combustor from Compressor Exit to Turbine Inlet,'' J. Eng. Gas Turbines Power, **121**, pp. 89–95.
[8] Spencer, A., 1998, ''Gas Turbine Combustor Port Flows,'' Ph.D. thesis, Loughborough University.
[9] Bicen, A. F., 1981, ''Refraction Correction For LDA Measurements in Flows With Curved Optical Boundaries,'' Imperial College, Fluids Section, Report FS/81/17.
[10] Launder, B. E., and Spalding, D. B., 1974, ''The Numerical Computation of Turbulent Flows,'' Comput. Methods Appl. Mech. Eng., **3**, pp. 269–289.
[11] Patankar, S. V., and Spalding, D. B., 1972, ''A Calculation Procedure for Heat, Mass and Momentum Transfer in Three-Dimensional Parabolic Flows,'' Int. J. Heat Mass Transf., **15**, pp. 1787.
[12] Ferziger, J. M., and Peric, M., 1996, *Computational Methods for Fluid Dynamics*, Springer-Verlag, Berlin.
[13] Thompson, J. F., Warsi, Z. U. A., and Mastin, C. W., 1985, *Numerical Grid Generation: Foundations and Applications*, North-Holland, New York.

M. De Lucia

R. De Sabato

Dipartimento di Energetica,
Università degli Studi di Firenze,
Florence, Italy

P. Nava

S. Cioncolini

Nuovo Pignone General Electric,
Florence, Italy

Temperature Measurements in a Heavy Duty Gas Turbine Using Radiation Thermometry Technique: Error Evaluation

This paper describes a technique for the analysis of temperature measurement data obtained by using a radiation thermometer in a heavy duty gas turbine (PGT2 from Nuovo Pignone GE). The main sources of error are shown and analyzed and their influence on measurements and techniques used to correct data read. Success in significant error elimination in acquisition results in a more realistic and correct description of target temperature distribution [DOI: 10.1115/1.1341205]

Introduction

Infrared temperature measurement techniques are used when contact with a hot target is not advisable or even impossible. This is the case, for example, with moving bodies, targets not achievable by traditional instruments, surfaces characterized by such a high temperature as to risk damaging contact sensors or when the same contact with the measuring instrument can modify the target temperature. The advantages in using a radiometric thermometer instead of the traditional contact sensor in gas turbine monitoring are as follows:

- a high level of working frequency (both for the real optic sensor and for the electronic conditioning system);
- a high sensitivity and accuracy of measuring temperature;
- the characteristic of not being an intrusive scanning system (to avoid thermal and fluidynamic interference in turbomachinery flows);
- the possibility, provided by a flexible optic fiber cable, of keeping all the electronic instruments far away from the working area.

On the other hand, there are some drawbacks:

- the need to work with very high optic efficiency;
- the limitation of using the radiation thermometer only in the temperature range for which it was designed;
- a more complex management of the system (the need, for an excellent spatial and temporal resolution, use of uncertainty parameters, limitation and/or control of optic contamination);
- higher cost of this system as compared with traditional ones;
- frequent need for a technician for obtaining efficient measurements.

Measurement Errors. Errors occurring in turbomachinery radiometric scanning are mainly due to the hostile environment in which the instrument must work and this causes the optic contamination of the signal reaching the infrared sensor. The main error sources are:

(a) signal dumping due to the inclination between target surface and radiation thermometer axis;
(b) emitted and reflected radiation coming from hot combustion gas constituting the turbine stage environment;
(c) interference due to transient events;
(d) emitted and reflected radiation coming from the mirror system used for optic signal deviation;
(e) in a GT environment, radiation reaching a sensor is due not only to that emitted by the target, due to its temperature, but also to that reflecting from the surrounding hot areas.

Influence of Target Inclination Angle. Working in perfect environmental conditions, a radiation thermometer can measure the target temperature correctly when its surface is exactly perpendicular to the instrument axis. A target angular inclination causes a signal reduction that, at a certain stage, becomes unacceptable for realistic scanning. So, it is important to determine the exact calibration curve, for each particular radiation thermometer used in measurements, like the one shown in Fig. 1.

As Fig. 1 shows, errors in measuring can be neglected only when working with an inclination lower than 30 deg. Then, it is really important to guarantee this condition or predispose a system able to determine this angle and make an automatic correction of the acquired signal.

Emitted and Reflected Radiation Coming From Environment Gas. Combustion gas present in the first stages of turbo engines consists of steam, CO_2, CO, and SO_2, unburned particles and dust. These constituents cause a considerable number of emission, reflection, and scattering phenomena that can be sources of serious temperature measurement errors.

The emission due to a monatomic gas concerns the UV wavelength and only a small area of the infrared region (the one nearest to the visible band).

Symmetrically biatomic molecules (O_2, N_2, etc.) are transparent to infrared rays at low and middle temperatures, and the same happens to the asymmetric ones (CO, NO, etc.) whose emission can be neglected in the radiation thermometer working field.

Triatomic molecules (H_2O, CO_2, SO_2, N_2O, etc.) are, on the other hand, characterized by a strong emission in the infrared region so that their influence must be carefully assessed. This problem could be solved by considering radiative behavior of combustion gases, their composition and temperature but, in the modern GT, concentration of these components can change strongly and rapidly from one area to another. So the best solution is to overcome this problem by using filters able to restrict the working wavelength range to the field in which the influence of hot gas is minimized ($\lambda = 1-1.2\ \mu$m).

Interference due to Transient Events. The main transient events in radiometric scanning of a gas turbine are due to the sight of the flame coming from the combustion chamber and to carbon particles in exhaust gas. The effect of complex transient events is generally not such a big problem in heavy duty gas turbines.

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Indianapolis, IN, 7–10 June 1999; ASME Paper 99-GT-311. Manuscript received by IGTI, March 1999; final revision received by the ASME Headquarters, August 2000. Associate Editor: H. D. Nelson.

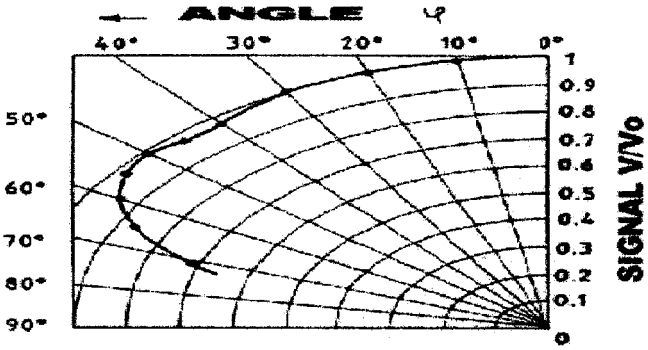


Fig. 1 Signal reduction due to inclination

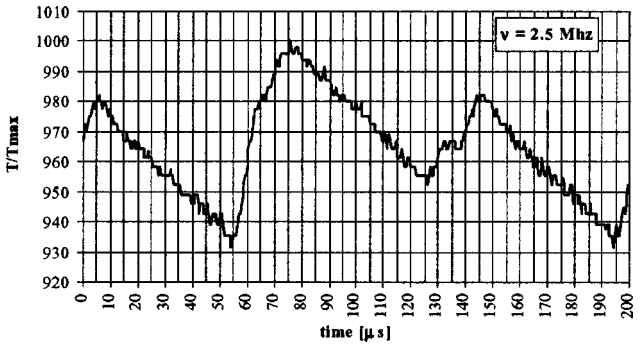


Fig. 3 Temperature measurement at 2.5 MHz (2MW-GT)

There is a problem, however, with particles that have not completed their combustion and are dragged from the flow in the first turbine stages. Generally, they are particles with a high volume-surface ratio that, while finishing their oxidation, are like small and strong moving radiation sources whose influence on measurement is significant since they cause temperature over-evaluation (shown by unreasonable and nonperiodic peaks in the output signal). This problem can be totally eliminated by simple data post-processing (i.e., filtering of nonperiodic spikes).

Radiation Emitted and Reflected by Mirror. The optic signal arriving from the target to the radiation thermometer photosensor can be direct or diverted by using one or more mirrors. This technique is used when it is necessary to increase the instrument view area to obtain a complete blade radiometric scanning. Of course, their presence causes some problems in terms of radiation emission and reflection due to the mirrors themselves.

The mirror must be cooled in order to reduce radiation emission as far as possible. Experimental results show that it is sufficient to cool the mirror at temperature at least 100°C less than the target one, to avoid all sensitive influence on measurements. There is also the problem of the constancy of optic mirror characteristics. The experiment, analyzed in following sections, was carried out without using the mirror system.

Experimental Setup. In order to acquire rotor blade temperature distribution, the radiation thermometer was positioned as shown in Fig. 2(*A*). Turbine rotation makes the target move along the blade from TIP to HUB till a point at which the sight radiation thermometer is covered by the next blade passage. By changing instrument inclination, it is possible to move this point as near to or as far away from the HUB as necessary. The instrumentation used was:

(a) radiation thermometer: model R7QB from Land Infrared customized;
(b) spot dimension: 4 mm circular with focal length of 100 mm;
(c) blade area inspected: 16 mm from the TIP (c. 60 percent of the blade).

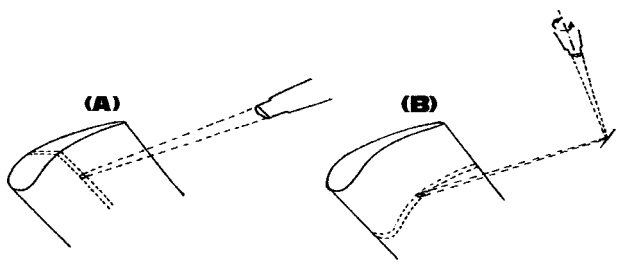


Fig. 2 Comparation between views without and with mirrors

Experiments were carried out trying to limit the optic contamination reaching the radiation thermometer sensor. These tests were repeated using different sampling rate values to evaluate the influence of a different spatial resolution. In Figs. 3 and 4 test results obtained on the first rotor stage with a sampling rate of 2.5 Mhz (the highest) and 0.1 Mhz (the lowest) are shown.

As can be seen, the use of a higher sampling rate allows a better description of temperature distribution due to the possibility of acquiring the smallest variation in time intervals. On the other hand, it also allows to get more details on the coming radiation. The temperature distribution obtained in this way cannot be considered as representative of the real situation. In fact:

- a turbine stage is a nonisothermal environment so that in target radiative temperature measurement it is necessary to consider the influence of the surroundings;
- the radiation thermometer works using a finite resolution that causes the loss of some information with the consequent signal reduction during scanning.

Radiation Reaching. In a nonisothermal environment, radiation measured by a radiation thermometer is due not only to that actually emitted by the target, due to its temperature, but also to that reflecting from the surrounding hot areas onto the target adding to that measured by radiation thermometer [1–5]. Scientific literature suggests considering the influence of reflected radiation by using a correction coefficient, named apparent emissivity (ε_a), that must be found, on each occasion, experimentally [6,7]:

$$\varepsilon_a = \frac{L_{ex}(\lambda,T)}{L_b(\lambda,T)} = \frac{L_{em}(\lambda,T)+L_r(\lambda,T)}{L_b(\lambda,T)}.$$

So as not to complicate the analysis, a value equal to one is often suggested. However, in this way what is measured is an altered blade temperature distribution, which is often over-valued. For this reason, the radiative model suggested by De Lucia and Lanfranchi in 1992 [8], and subsequently upgraded by De Lucia and Masotti in 1994 [9], was preferred. This model uses the calcula-

Fig. 4 Temperature measurement at 0.1 MHz (2MW-GT)

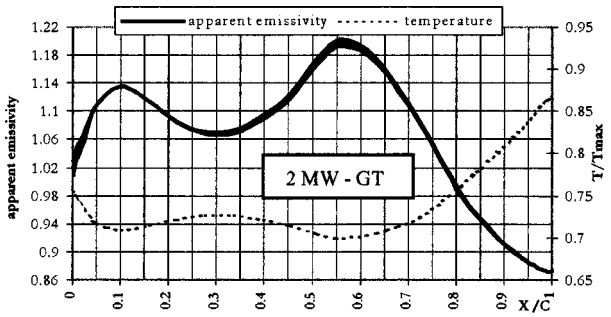


Fig. 5 Apparent emissivity band (rotor blade)

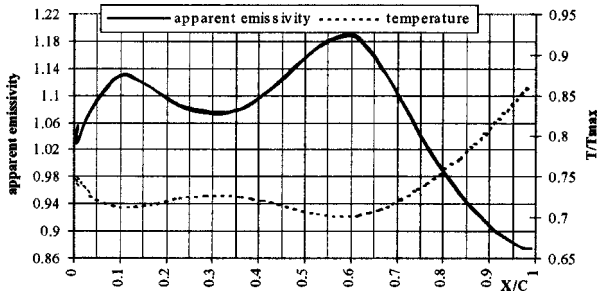


Fig. 7 Apparent emissivity for fictitious GT

tion of view factors between the different surfaces in which the geometric problem is resolved and allows exact calculation of the value of apparent emissivity for each blade point. In this way, it is possible to process the outgoing target radiation, the one measured by the sensor, in order to determine the radiation that a black body would emit at the same target temperature. By inverting Planck's law [10,11], it is then possible to evaluate the real target temperature.

In the following, Fig. 5 apparent emissivity distribution is shown, with regard to the pressure side of a first turbine stage rotor blade, and the consequent dimensionless temperature distribution against the target position on the blade (seen as a percentage of the axial chord).

It can be seen that the apparent emissivity distribution is approximately opposed to the temperature distribution so that the higher the temperature the lower the apparent emissivity determined at the same point.

Apparent emissivity values, for first rotor stage, range from a minimum of 0.87, on the suction side, to a maximum of 1.2 on the pressure side (from 0.8 to 2 for the stator blade). So, the usual hypothesis of considering a value of 1 becomes an acceptable approximation only when a high level of precision for measurements is unnecessary since errors on ϵ_a evaluation can reach values of 20 percent (even 100 percent for the stator blade).

We also need to consider that apparent emissivity is a function of temperature distribution but also of geometric disposition of all the target surfaces. To consider the influence of geometry, the calculation was repeated in ten different positions between rotor and stator. This continuous changing of the geometry due to the turbine rotation causes a periodic variation of apparent emissivity calculated. So, this is a function of time as well as of spatial position. However, looking at the narrow band (see Fig. 5) in which the apparent emissivity values are positioned, it is possible to realized there is not a big error in considering a mean value of the ten positions.

In terms of temperature, the exact calculation of apparent emissivity, instead of using a value of one, allows avoidance of overevaluation of temperature distribution (15°C–45°C) far too high to be accepted.

To confirm these results, the same analysis was carried out for a new generation turbine characterized by a geometry and a cooling system very different from the previously seen GT. Even in this case (see Fig. 6), the apparent emissivity distribution is opposite to the temperature distribution:

And here the temperature error due to considering the mean value of ε_a instead of its temporal one is minute (less than 0.3 percent versus 0.2 percent of the previous case).

Influence of Geometry and Temperature Distribution on Determining of ε_a

To understand better the importance of temperature distribution and system geometry with regard to apparent emissivity calculation, the same process shown before was repeated for a fictitious turbine characterized by the same geometry of that of a latest generation machine (5 MW-GT) and the temperature distribution of the first one examined in this paper (2 MW-GT). The results obtained (see Fig. 7) are only indicative and cannot demonstrate anything but are of considerable interest.

They show that apparent emissivity distribution is similar to the one just found for the rotor blade of the first turbomachine analyzed (see Fig. 5). This means that theoretical temperature distribution is the main parameter influencing apparent emissivity value while much less remarkable is the effect of view factors (and then the influence of geometry).

The Problem of Finite Resolution

During data aquisition, the radiation thermometer can read temperature values along the blade side due to its rotation movement. The view is only possible up to the moment at which the TIP of the following blade covers the previous one (see Fig. 8).

At the passage between a blade and the following one, when the radiation thermometer sees two very different temperature distri-

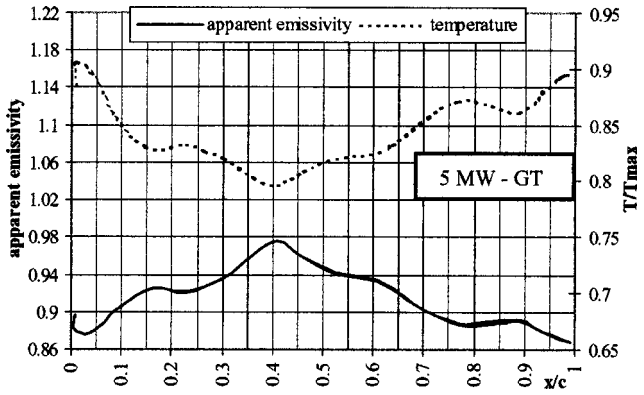


Fig. 6 Apparent emissivity (rotor blade)

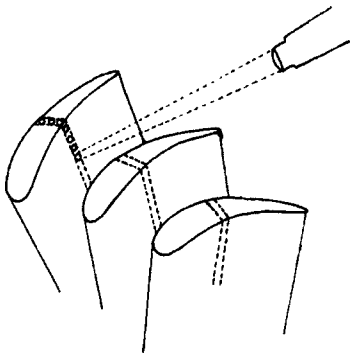


Fig. 8 Scanning blade measurement (along blade height)

bution (high temperature for the blade TIP and low temperature for HUB), there is a major error in aquisition. The radiation thermometer spot (the area seen by instrument focal system) has a finite dimension (it is not really a point as theory supposes) and a particular shape. The consequence is that the temperature acquired is not the real one relative to the point seen by the instrument but an average value of the temperature of all the blade points in the radiation thermometer spot area. So, during blade passage, the measured temperature is a value that is not representative of any real situation.

To understand the influence of spot shape, it is possible to see the variation of temperature measurement, obtained by a radiation thermometer that scans an isothermal surface, against the percentage of area seen by the instrument when a rectangular or a circular spot is used. Results are reported in dimensionless terms in Fig. 9.

As can be seen, rectangular spot behavior is linear while that of the circular spot is different, and more complex (the latter is the most common). Apart from spot shape, there are two other really important parameters: the spot dimension and the signal rate sampling that cause an alteration of measurement during blade passage. The problem of correction of this kind of error was faced by considering a sawtooth signal and simulating its scanning by radiation thermometer. The main parameters are the value of LR (the ratio between the blade visible length 1 and the radius R_{spot} of the circular spot) considering the geometric dimensions of the problem and the value of BW (ratio between the rate sampling and the machinary rotation frequency) considering the accuracy by which the signal is scanned. Naming the abscissa determining the spot position on blade z, the dimensionless equations characterizing instrument behavior (circular spot) in the regions in which there are the three different temperature distributions are as follows:

$$0 \leqslant z \leqslant R_{spot}$$

$$\frac{T_{max}(z)}{T_{max}} = \left(1 - \frac{z}{l}\right) + \frac{1}{\pi} \cdot \left(\frac{z}{R_{spot}}\right) \cdot \sqrt{1 - \left(\frac{z}{R_{spot}}\right)^2} + \frac{1}{\pi} \arcsin\left(\frac{z}{R_{spot}}\right) - \frac{1}{2}, \tag{1}$$

$$R_{spot} \leqslant z \leqslant l - R_{spot}, \quad \frac{T_{meas}(z)}{T_{max}} = 1 - \frac{z}{l}, \tag{2}$$

$$l - R_{spot} \leqslant z \leqslant l$$

$$\frac{T_{meas}(z)}{T_{max}} = \left(1 - \frac{z}{l}\right) + \frac{1}{2} - \frac{1}{\pi} \cdot \left(\frac{l}{R_{spot}} - \frac{z}{R_{spot}}\right) \cdot \sqrt{1 - \left(\frac{l}{R_{spot}} - \frac{z}{R_{spot}}\right)^2} - \frac{1}{\pi} \cdot \arcsin\left(\frac{l}{R_{spot}} - \frac{z}{R_{spot}}\right). \tag{3}$$

Since the measurement of z abscissa cannot be carried out continuously but in an interrupted manner, it is possible to write the discrete space covered at instant i as $z_i = z_{i1} + dz$, $dz = \text{L/BW}$ being the incremental spatial step made by the blade (and measured in a direction perpendicular to that of the radiation thermometer axis) between two following readings. Now it is possible to write the following ratios:

$$\frac{l}{R_{spot}} = \text{LR}; \quad \frac{z}{l} = \frac{i \cdot dz}{l} = i \cdot \frac{1}{\text{BW}};$$

$$\frac{z}{R_{spot}} = \frac{i \cdot dz}{R_{spot}} = \cdot \frac{i \cdot dz}{L} \cdot \frac{l}{R_{spot}} = i \cdot \frac{1}{\text{BW}} \cdot \text{LR}.$$

In this way, all the previous expressions can be written as a function only of LR and BW. Then, changing the value of these two ratios, two curve families were generated to show their real influence in data aquisition.

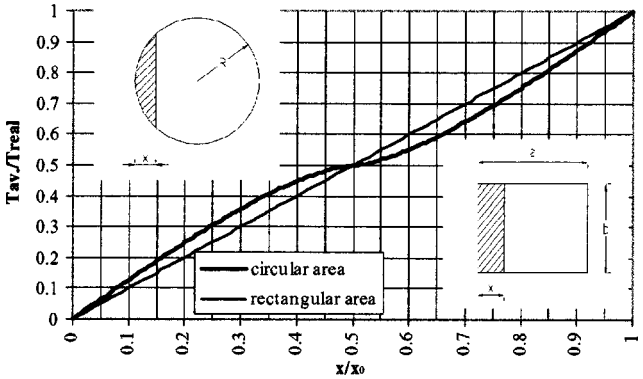


Fig. 9 Influence of spot shape

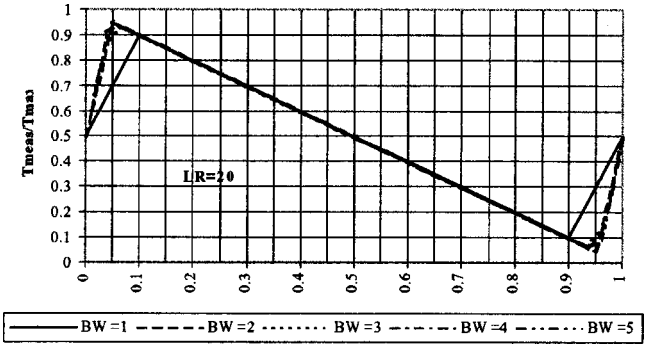


Fig. 10 Signal reduction for circular spot varying BW

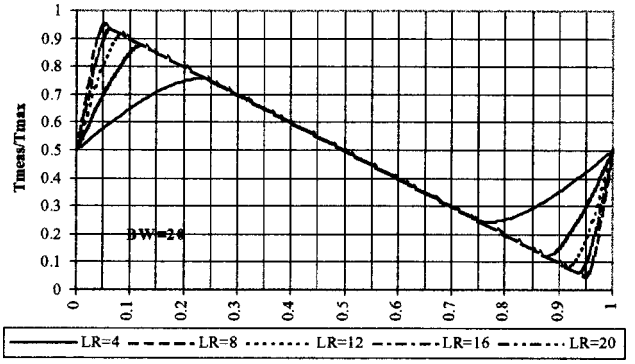


Fig. 11 Signal reduction for circular spot varying LR

As can be seen in Fig. 10, for a fixed value of LR, the greater the sampling rate, in respect of rotation frequency, the lower the reduction of the signal measured in respect of the real one. On the other hand, the lower the sampling rate the greater the loss of information so that the accuracy in describing temperature distribution is lower (it is possible to have a reduction even greater than 10 percent; see Fig. 10 BW=1).

In our tests, sampling rate changed from 0.1 Mhz to 2.5 Mhz while the rotation speed was 22,500 rpm. So, BW changes from 266 to 6667 with a consequent negligible signal reduction.

The LR ratio, in our case, has a value of 10 so that, considering the worst values for BW, it is even possible to have a reduction greater than 10 percent, as is shown in Fig. 11.

Data Analyzing Model

Temperature measurement is the result of an integration on spot area of the radiation really emitted by the target. Information loss concerning temperature distribution in spaces smaller than the

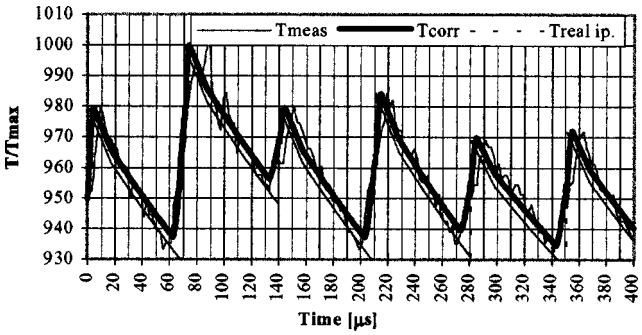


Fig. 12 Simple sawtooth model

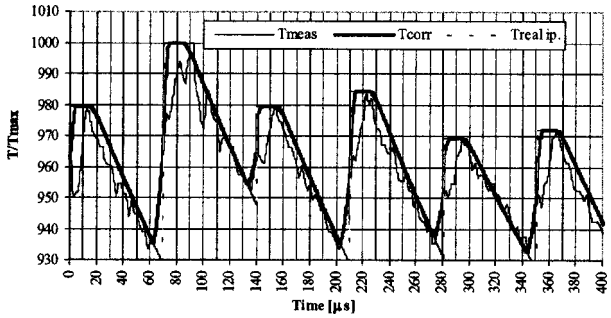


Fig. 13 Trapezoidal model

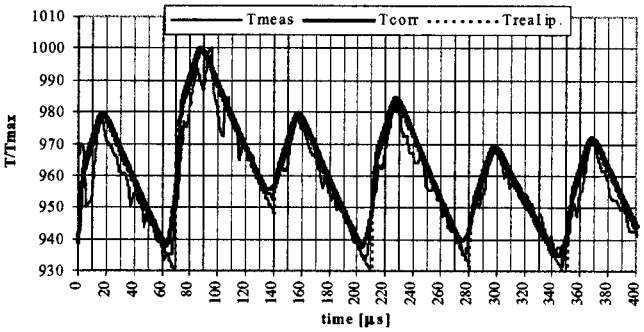


Fig. 14 Modified trapezoidal model

spot diameter causes the impossibility of rebuilding the real signal on the basis of acquired data. To overcome this problem we built up a theoretical model of temperature distribution, simulating realistically its scanning (same values for LR and BW) to obtain a corrupted signal and compare it with the signal acquired by using a radiation thermometer on the turbomachinery. The greater or lesser coincidence between the corrupted theoretical signal and the real measurement gives some information about the goodness of the initial theoretical model. Since incorrect behavior of the cooling system of a blade, superficial irregularity, or local oxidation of some part of another blade can justify a different temperature distribution between blades, obtaining any signal periodicity, i.e., an identical distribution on every blade, cannot be contemplated. In this way, the theoretical model must be built up starting from the signal measured and using an iterative model to search for the best congruence between the corrupted and measured ones.

Many theoretical models were built up. Those the authors consider the best for analysis are presented here:

(a) simple sawtooth model (see Fig. 12);
(b) trapezoidal model (see Fig. 13);
(c) modified trapezoidal model (see Fig. 14).

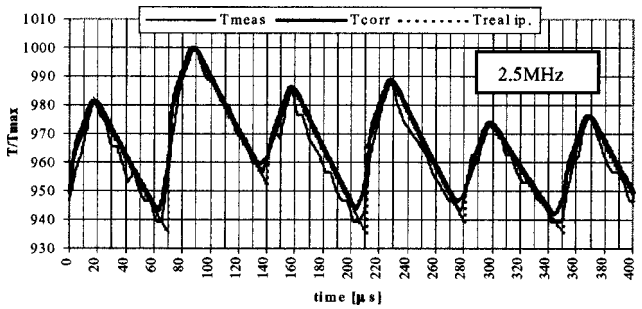


Fig. 15 2MW GT temperature measurement

In Fig. 15, the application can be seen of each of these models to the same temperature data series. In the following graphs, the experimental measured signal (T_{meas}), the theoretical model adopted ($T_{\text{real ip.}}$), and the corrupted one (T_{corr} obtained integrating the theoretical signal on the spot area) are shown.

The first model seems to fit well the acquired signal; the correspondent RMS value, that is the evaluation parameter adopted to find the best theoretical model is 0.085. The second case seems to be the worst; here is an increase of RMS that reaches a value of 0.104. The last model is closest to the signal measured; RMS value is 0.06, the lowest, so that this is the model used for data analysis.

Experimental Results

Analysis was carried out on different data series obtaining in each case coherent and comparable results. Thus only one of them is shown here.

Difference between corrupted signal and acquired one is always less than 1 percent.

After defining the model, the error due to the finite resolution with which the radiation thermometer works can be estimated.

As shown in Fig. 16, errors become noticeable only during blade passage. In terms of temperature, the maximum error is about 2 percent (only in few cases is it more than 3 percent).

Error Evaluation

The overall error without any correction can be evaluated as follows:

• First correction seeks to avoid the effect of finite resolution in temperature measurement. Naming $T_{\text{fin.res.}}$ the temperature obtained on the basis of this correction and $e_{\text{fin.res.}}$ the percentage difference between $T_{\text{fin.res.}}$ and the measured temperature, it is possible to write this expression:

$$e_{\text{fin.res.}} = \frac{T_{\text{fin.res.}} - T_{\text{meas}}}{T_{\text{fin.res.}}} \Rightarrow T_{\text{fin.res.}} = \frac{T_{\text{meas}}}{(1 - e_{\text{fin.res.}})}.$$

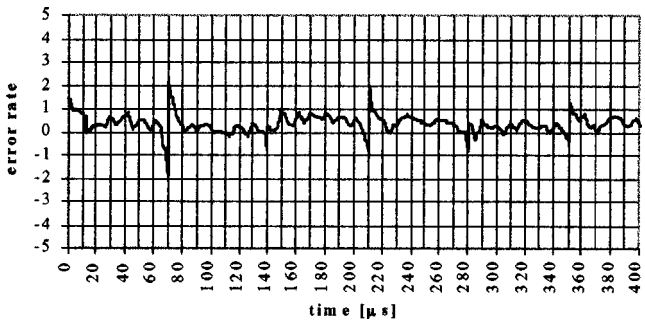


Fig. 16 Temperature error ($T_{\text{meas}} - T_{\text{corr}}$)

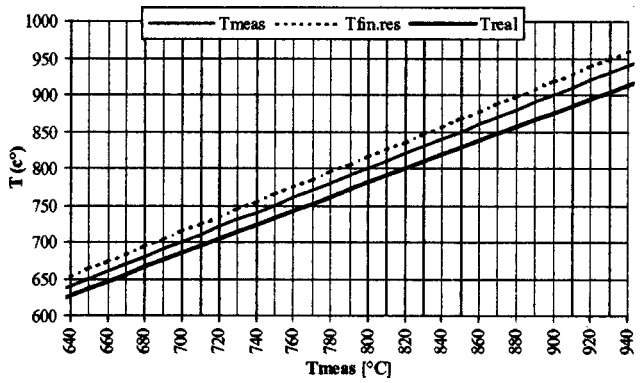


Fig. 17 Temperature correction vs T_{meas}

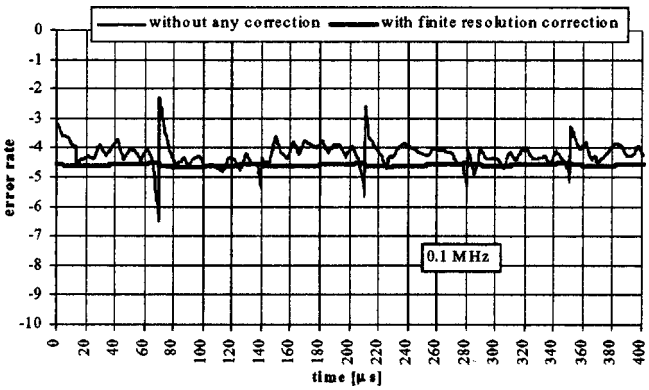


Fig. 19 Percentage error rate at 0.1 MHz

• Then the correction due to apparent emissivity must be introduced in order to consider the not so ideal environment. Defining the intermediate temperature as radiance temperature T_λ obtained when errors due to the physics phenomena and imperfections in radiation thermometer working are eliminated (in this case $T_\lambda = T_{\text{fin.res.}}$), we can define a new error:

$$e_{\text{real}} = \frac{T_{\text{real}} - T_\lambda}{T_{\text{real}}}$$

so that we have

$$T_{\text{real}} = \frac{T_\lambda}{1 - e_{\text{rad}}} = \frac{T_{\text{meas}}}{(1 - e_{\text{rad}}) \cdot (1 - e_{\text{fin.res.}})}.$$

Temperature distribution against the generic T_{meas} measured, without considering any correction (T_{meas}), considering the finite resolution of radiation thermometer ($T_{\text{fin.res.}}$) and in the case in which we consider even the effect of apparent emissivity (T_{real}), are all plotted in Fig. 17.

In terms of temperature, these errors correspond to ΔT of 30°C–50°C that means an error up to 6 percent.

As can be seen, these two corrections have opposite effect: the first one causes a temperature distribution increase because of the elimination of reduction due to finite resolution; on the other hand, the second causes a decrease in temperature, owing to the elimination of signal over-evaluation due to the contribution of apparent emissivity (reflected radiation coming from surroundings).

Now it is possible to see, for data obtained by sampling with the maximum rate (2.5 Mhz), which is the real error without considering any correction or considering only the effect of finite resolution.

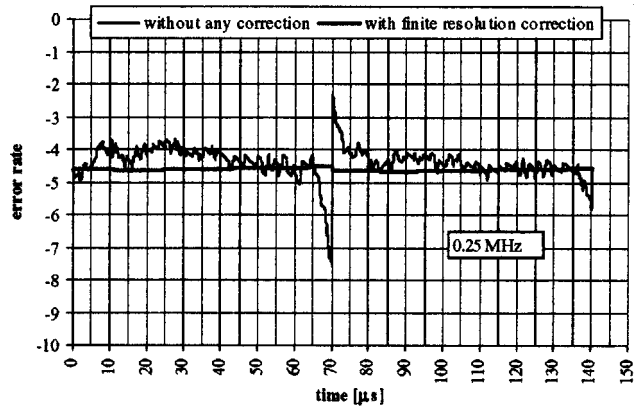


Fig. 18 Percentage error rate at 2.5 MHz

As can be seen in Fig. 18, not using any correction causes an error whose peaks are even greater than 7 percent (during blade passage) and whose average value is about 4 percent–5 percent.

The same operation can be repeated for the lower sampling rate (0.1 Mhz); see Fig. 19.

Here, again, not using any temperature correction causes an error whose peaks occur during blade passage while their average value is about 4 percent–5 percent.

Conclusions

An extension of the data correction model would foresee the improvement of a tridimensional model to better determine view factors (the model used is bidimensional) to consider the ever more complex geometry of modern turbomachinery. This way of working seems to be far better than that followed and proposed by scientific literature up to now (in which corrections are absent or are extremely approximate, e.g., considering $\varepsilon_a = 1$).

In this analysis, correction due to surface inclination is not considered. So there is another error whose value can reach about 3 percent–4 percent.

However, by using the proposed correction it is possible to avoid errors, in temperature evaluation, whose average value for working temperature up to 950°C is about 40°C with peaks of 65°C, in this way allowing better system monitoring and a more realistic description of blade temperature distribution.

Future developments could regard the definition of a better theoretical model in order to improve fitting to obtain the best possible final result.

Acknowledgments

The authors are grateful to the NuovoPignone experimental staff for their help in carrying out the experimental program and to Professor Ennio Carnevale for his invaluable support.

Nomenclature

BW = sampling rate to machinery rotation ratio
D, R = diameter and radius
E = error
L = radiance W m^{-2} sr^{-1} μm^{-1}
l = visible blade length
LR = blade visible length to radius spot ratio
T = temperature
z = abscissa of spot position on blade
ε = emissivity
λ = wavelength μm

Subscripts

a = apparent
av = average
b = blackbody

corr = corrupted
em = emitted due the temperature surface
ex = exitent
fin.res = finite resolution
meas = measured
r = reflected
real.hp = theoretical hypothesis

References

[1] Beynon, T. G. R., 1981, ''Turbine Pyrometry-An Equipment Manufacturer's Viewpoint,'' ASME Paper 81-GT-136.

[2] Beynon, T. G. R., 1982, ''Infrared Radiation Thermometry Applied to the Development and Control of Gas Turbine Engines,'' *Proceedings, International Conference on Infrared Technology*, Butterworth Scientific, Ltd., Sevenoaks, U.K., pp. 3–16.

[3] DeWitt, P. D., and Incropera, F. P., 1988, ''Physics of Thermal Radiation,'' *Theory and Practice of Radiation Thermometry*, edited by P. D. DeWitt and G. Nutter, Wiley, New York, NY, pp. 21–89.

[4] Douglas, J., 1980, ''High-Speed Turbine Pyrometry in Extreme Environments,'' Proceedings, New Orleans Gas Turbine Conference, ASME, New York, NY, pp. 335–343.

[5] Kirby, P. J., Zachary, R. E., and Ruiz, F., 1986, ''Infrared Thermometry for Control and Monitoring of Industrial Gas Turbines,'' Proceedings, International Gas Turbine Conference and Exhibit, Duesseldorf, Germany, ASME Paper 86-GT-267.

[6] Schulemberg, T., and Bals, H., 1987, ''Blade Temperature Measurements of Model V84.2 100 MW/60 Hz Gas Turbine,'' Proceedings, International Gas Turbine Conference and Exhibit, Anaheim, CA, ASME Paper 87-GT-135.

[7] Scotto, M. J., and Eismeier, M. E., 1980, ''High-Speed Noncontacting Instrumentation for Jet Engine Testing,'' ASME J. Eng. Power, **102**, pp. 912–917.

[8] De Lucia, M., and Lanfranchi, C., 1992, ''An Infrared Pyrometry System for Monitoring Gas Turbine Blades: Development of a Computer Model and Experimental Results,'' 1992, IGTI-92, Paper No. 92-GT-80. ASME, Journal of Engineering for Gas Turbines and Power, January 1994, Vol. 116, pp. 172–177.

[9] DeLucia, M., and Masotti, G., 1994, ''A Multimeasurement Infrared Pyrometry System for Determining Temperature Distribution in Gas Turbines,'' IGTI 94, Paper No. 94-GT-39, ASME, Journal of Engineering for Gas Turbines and Power, April 1995, Vol. 117, pp. 341–346.

[10] Love, T. J., 1988, ''Environmental Effects on Radiation Thermometry,'' *Theory and Practice of Radiation Thermometry*, edited by P. D. DeWitt and G. Nutter, Wiley, New York, NY, pp. 189–229.

[11] Ono, A., 1988, ''Methods for Reducing Emissivity Effects,'' *Theory and Practice of Radiation Thermometry*, edited by P. D. DeWitt and G. Nutter, Wiley, New York, NY, pp. 565–623.

P.-J. Lu
Institute of Aeronautics and Astronautics,
National Cheng Kung University,
Tainan, Taiwan
pjlu@mail.ncku.edu.tw

M.-C. Zhang
Department of Jet Propulsion and Power,
Beijing University of Aeronautics
and Astronautics,
Beijing, China

T.-C. Hsu
Institute of Aeronautics and Astronautics,
National Cheng Kung University,
Tainan, Taiwan
tjshyu@mail.ncku.edu.tw

J. Zhang
Department of Jet Propulsion and Power,
Beijing University of Aeronautics
and Astronautics,
Beijing, China
cdq-rfs@263.net

An Evaluation of Engine Faults Diagnostics Using Artificial Neural Networks

Application of artificial neural network (ANN)-based method to perform engine condition monitoring and fault diagnosis is evaluated. Back-propagation, feedforward neural nets are employed for constructing engine diagnostic networks. Noise-contained training and testing data are generated using an influence coefficient matrix and the data scatters. The results indicate that under high-level noise conditions ANN fault diagnosis can only achieve a 50–60 percent success rate. For situations where sensor scatters are comparable to those of the normal engine operation, the success rates for both four-input and eight-input ANN diagnoses achieve high scores which satisfy the minimum 90 percent requirement. It is surprising to find that the success rate of the four-input diagnosis is almost as good as that of the eight-input. Although the ANN-based method possesses certain capability in resisting the influence of input noise, it is found that a preprocessor that can perform sensor data validation is of paramount importance. Autoassociative neural network (AANN) is introduced to reduce the noise level contained. It is shown that the noise can be greatly filtered to result in a higher success rate of diagnosis. This AANN data validation preprocessor can also serve as an instant trend detector which greatly improves the current smoothing methods in trend detection. It is concluded that ANN-based fault diagnostic method is of great potential for future use. However, further investigations using actual engine data have to be done to validate the present findings.
[DOI: 10.1115/1.1362667]

Keywords: Engine Condition Monitoring, Fault Isolation, Artificial Neural Networks

Introduction

Model-based engine condition monitoring approach (Urban, 1972 [1]; Urban and Vopolni, 1992 [2]; Doel, 1994 [3,4]) has not yet been generally accepted as a routinely used tool in the practical applications. The engine condition monitoring (ECM) systems currently in use in the airlines can be divided into two categories, namely, the limited and the expanded systems. The limited (three- or four-input parameters) system suffers from the unidentifiability problem, making performance trending and fault diagnosis rely mainly on the checkup work using limited available fingerprint plots. Although strict Kalman filter-type analysis can be performed for the expanded (eight or more input parameters) system, the analysis calls for specially trained experts who usually are not available in airlines.

Neural network-based approach circumvents all the above fundamental difficulties and emerges as a potential tool to carry out engine condition monitoring and modular fault isolation. The use of artificial neural network (ANN) approach (Zedda and Singh, 1998 [5]; Whitehead et al., 1990 [6]; Eustace and Merrington, 1995 [7]; Roemer and Atkinson, 1997 [8]; Cifald and Chokani, 1998 [9]) for performance trending and fault diagnosis could be simple and efficient provided the network is properly trained. This advantage allows performance analysis be automated and, as pursued further, can in the future make condition monitoring a daily or even flight-to-flight routine. Safety engine operation and effective on-condition maintenance can thus be guaranteed on a firmer basis. The present work aims at exploring the ability and limitations associated with the ANN-based engine fault diagnostic approach. These findings will be valuable as real engine data are to be used as the input in the future.

Training and testing data are supplied by Pratt and Whitney (P&W) using influence coefficient matrix pertaining to PW4000 engines. The standard deviations obtained from various data retrieving methods are used to model the data scatter. Simulated data using this noise model and the influence coefficient matrix are generated. Back-propagation (BP) neural networks Werbos, 1974 [10]; Haykin, 1994 [11]) are trained and tested using the simulated data. The performance of both four-input and eight-input BP diagnostic networks is evaluated. The influence of data scatter on the diagnostic success rate is carefully assessed. It will be demonstrated in the present investigation that both four-input and eight-input BP networks can achieve high success rates, exceeding the minimum 90 percent requirement, in diagnosing the selected 15 engine faults. Although ANN-based fault diagnostic methods have certain capabilities in resisting data contamination while making correct diagnostic judgment, a sensor data validation preprocessor is still considered important for the present method to be useful in the practical application.

To improve the input data quality, a preliminary autoassociative neural network (AANN) analysis is proposed for noise filtering (Romer, 1998 [12]; Kramer, 1991 [13], 1992 [14]; Napolitano et al., 1996 [15]; Mattern et al., 1997 [16], 1998 [17]). It is shown that significant improvement can be achieved when autoassociative filter is introduced to augment the fault diagnostic BP network. In addition, this AANN can serve as a trend detector which provides instantaneous judgement on whether true trend change occurs. Finally, concluding remarks are drawn and recommendations for future work are given, suggesting a roadmap for the use of engine data of the present ANN-based approach.

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Munich, Germany, May 8–11, 2000; Paper 00-GT-029. Manuscript received by IGTI Feb. 2000; final revision received by ASME Headquarters Jan. 2001. Associate Editor: M. Magnolet.

Back-Propagation Artificial Neural Network

The back-propagation (BP) algorithm is adopted herein for constructing the multilayer perceptrons which constitute the neural network used for fault diagnosis. The BP network is categorized as a supervised learning process, in which information sensed is propagated forward while error, after corrected against some targeted value, are transmitted backward through the network. For researches conducted in the area of ANN-based engine fault isolation, the BP algorithm has been demonstrated as the most popular and successful method.

The BP network used in the present work consists of three layers, namely the input, hidden, and output layers. The number of input nodes could be four or eight, depending on the number of sensors available. EGT, WF, N1, and N2 are assigned for the four-input network and the additional parameters, P25, T25, T3, and P3 are added onto the four-input ones to result in the eight-input network. As for the output nodes, at most 23 types of faults including engine and sensor indication errors are incorporated. The type and explanation of these faults are listed in Table 1 below. The abbreviation for each fault is also listed in the right column of Table 1.

In training the BP network, the synaptic weights w_{ij} are corrected using the following algorithm:

$$\Delta w_{ij}(n) = \eta\left(-\frac{\partial E(n)}{\partial w_{ij}}\right) + \alpha \Delta w_{ij}(n-1) \tag{1}$$

in which E is the targeted error, η and α are the learning rate and momentum constant, respectively. In the following applications, both η and α are kept within [0,1] and the appropriate values are determined by trial-and-error method.

Table 1 Analyzed faults

	Description	Type of Fault	Abbreviation
1	HPC Performance Loss	Type 1	HPC
2	HPT Performance Loss	Type 1	HPT
3	Stability Bleed (B25) Open	Type 1	25BS
4	Fan Discharge Area Change	Type 1	FP14
5	Primary Area Change	Type 1	FP8
6	LPC Performance Loss	Type 1	LPC
7	LPT Performance Loss	Type 1	LPT
8	Fan Performance Loss	Type 1	FAN
9	One Start Bleed (295) Open	Type 1	29B1
10	TCC System Full Fault	Type 1	TCC
11	Low (2.5) Bleed Leak	Type 1	25BL
12	High (2.9) Bleed Leak	Type 1	29BH
13	Both Start Bleeds (B2.95) Open	Type 1	29B2
14	HPC Stator Vane Misrigging	Type 1	SVM
15	P49 (EPR) Indication Problem	Type 2	+P49
16	N2 Indication Problem	Type 2	+N2
17	N1 Indication Problem	Type 2	+N1
18	EGT Indication Problem	Type 2	+EGT
19	WF Indication Problem	Type 2	+WF
20	P2 (EPR) Indication Problem	Type 2	+P2
21	TAT Indication Problem	Type 2	+TAT
22	Mach No. Indication Problem	Type 2	+Ma
23	Altitude Indication Problem	Type 2	+ALT

Type 1: Engine Fault Type 2: Instrumentation Error

Training and Testing Fault Samples. The quality of the training data holds a central position for the ANN diagnosis to be successful. For the present engine problem the input parameters are the deltas defined as the percentage difference of the measured data to its baseline value. Normally these deltas are on the order of a few percent which can easily be contaminated by random data scatter. How to make a correct diagnostic judgment under noisy or even incomplete input data situation becomes the main issue to be addressed in this work.

Fault samples to be used in the subsequent analysis are generated using PW4000-94 influence coefficient matrix (EPR=1.29), which connects the measurements with the performance related parameters, and the standard deviations of the sensor scatters. Noise-free data was generated by multiplying the influence coefficient matrix with some predetermined severity indicators pertaining to certain fault type. The final so-called simulated data, however, result from the addition of a noise model to the noise-free data. The formula is stated below:

$$sensor\ data = clean\ data + K\sigma[rand - 0.5] \tag{2}$$

where K is the control parameter governing the noise level, σ the characteristic standard deviation of the sensor scatter described in Table 2, and *rand* the random number between [0,1]. Five typical sensor data scatters listed in Table 2 are supplied by different user communities of airlines and engine repair stations. These data scatters are associated with unsmoothed raw data.

In order to result in an unbiased weight of importance for each output node in the supervised learning process, the sensor data in Eq. (2) is first normalized by its maximum value and then divided through by its scatter σ. This normalization can provide less-noise-contaminated data with relatively higher weight of influence.

Fault Diagnosis Using Simulated Engine Data

Effect of Noise Level. A thorough investigation was conducted using the P&W supplied data. Both four-input and eight-input ANN architectures were studied for diagnosing the faults listed in Table 1. Effort was exercised in finding the optimal selection of training/testing data set and the ANN network that may lead to the goal of 90 percent-or-higher success rate. Regardless of the input node number, the training success rate cannot exceed 95 percent and the averaged testing success rate can only achieve 60 percent or so. The unbalanced training and testing success rates indicates that the features of the fault patterns learned are inconsistent with the patterns provided in the test samples. In other words, the net is overtrained and the ability of generalization is lost. This is in sharp contrast to the results of ANN diagnosis using noise-free data, for which nearly 100 percent success rate can be achieved for both training and testing purposes. The cause for this low success rate is attributed to the high noise level assigned in generating the simulation data, as explained below.

Table 2 Standard deviations used for simulated engine data

Parameters	Case 1	Case 2	Case 3	Case 4	Case 5	Remarks
T49C2(deg. C)	7.26	8.30	3.34	3.81	4.30	+EGT deviation
WF(%)	1.09	0.91	0.72	0.74	0.78	Corrected +WF deviation
N2C2(%)	0.33	0.38	0.17	0.15	0.17	Corrected +N2 deviation
N1C2(%)	0.34	0.50	0.21	0.17	0.18	Corrected +N1 deviation
P25C2(%)	0.53	0.43	0.79	0.74	0.78	LPC pressure ratio deviation
T25C2(deg. C)	0.88	0.86	1.13	2.60	2.23	LPC exit temperature deviation
T3C2(deg. C)	13.02	5.42	2.16	3.25	2.96	HPC exit temperature deviation
P3C2(%)	0.70	1.05	0.45	0.84	1.00	HPC pressure ratio deviation

Table 3 Noise effect on simulated engine data

Parameters	T49C2	WF	N2C2	N1C2	P25C2	T25C2	T3C2	P3C2	Remarks
HPT ETA –1.09%	11.86	1.41	-0.62	0.082	1.41	1.24	-4.39	-1.37	Noise-free
HPT ETA –1.09%	2.65	1.06	-0.57	-0.63	0.39	0.33	1.85	-0.94	Noisy
Ratio	4.48	1.3	1.09	-0.13	3.62	3.76	-2.37	1.46	

A representative case illustrating the high scatter input data can be found with the HPT efficiency (ETA) loss, among others. Table 3 lists a comparison of noise-free and noisy data of a 1.09 percent HPT efficiency loss. Feature discrepancy can be found by comparing these two sets of data. It is observed that the ratio of each pair of parameter change differs far from the theoretical value of unity. This implies that the fault pattern corresponding to −1.09 percent HPT efficiency loss is hard to recognize as noise effect is included. For ANN-based diagnostic method, noise level effect is often the main cause responsible for difficulties in training the network and the low success rate in diagnosing testing samples.

Noise Level of Actual Engine Parameters. In order to verify what should be the proper magnitude of standard deviations, 48 PW4000 engines maintained by AMECO were analyzed. The actual engine data was extracted using the data extraction utility provided by P&W, which has a builtin ten-point moving average data smoothing capability. Figure 1 shows the raw and smoothed data and the five standard deviations of Table 2. It can be observed that the averaged standard deviations of raw data are comparable to cases 3, 4, and 5 of Table 2. The smoothed data, however, are significantly lower than those raw data given in Fig. 1, as shown in the statistical averages listed in Table 4.

During the course of examining AMECO data, it was found that some scatters due to trend changes are as high as the data scatters of cases 1 and 2 of Table 2. It seems the data jump due to trend change were accounted as data uncertainty in cases 1 and 2.

It is worth noticing that data used for diagnosis are smoothed data. Hence a reasonable suggestion for constructing simulation data that reflects the randomness of the measured engine data would be using standard deviations equivalent to cases 3, 4 and 5 together with choosing $K=2.0$.

Feasibility Study of ANN-based Fault Diagnostic Method Training and testing fault samples used for evaluating the ANN-based diagnostic method are generated using the scatter of case 5 and $K=2.0$. Since HPT and TCC faults, and +Ma and +ALT indication errors are indistinguishable as viewed from the simulated data, we combine these four fault types into two, resulting in a total of 21 faults that are to be diagnosed.

The BP network designed has a three-layer 4-15-21 node architecture. In the generation of training data, there are 50 samples generated for each fault. As for the testing data, the number of samples for each fault is increased up to 100. In other words, there will be 2200 cases to be diagnosed by a BP network which is trained using 1100 fault samples.

The test results are illustrated in Table 5. The overall success rate achieves a value of 91 percent which satisfies the original 90-or-higher design goal. Of the individual success rate for each fault, it is found that LPC, +EGT, +N2, and SVM are faults that are relatively difficult to identify. LPC fault is often confused with +EGT fault and +N2 and SVM faults are easily mixed up. The reason for this has been verified by changing the standard deviations of Eq. (2). The success rates are found inversely proportional to the level of data scatter, and LPC, +EGT, +N2, and SVM are faults that are more susceptible to noise contamination.

Evaluation of ANN-based Fault Diagnostic Method. The feasibility study of ANN-based diagnosis concluded that a minimum 90 percent success rate can be achieved for practical engine operations. Since P&W has already had a method to separate aircraft and engine instrumentation problems from the gaspath faults, it was agreed that the presently developed neural network focuses on the isolation of the first 15 faults listed in Table 1. In addition, the level of data scatter should be equivalent to those of cases 3, 4, and 5 of Table 2. Owing to the importance of data scatter on the accuracy of the present diagnostic method, a parallel research on noise filtering of the input data was conducted. The method and results are presented in the next section.

Both four-input and eight-input BP neural nets were constructed. For each fault, 20 and 50 cases are generated for training and testing purposes, respectively. Noise model adopts case 3 standard deviations and the noise control parameter K was selected from 1 to 3.5, covering a full spectrum of diagnosis under low to high noise levels. The threshold for the output layer to be fired is set to be 0.3, and a successful diagnosis is regarded as one of the fired three possible output predictions matches with the true fault.

The four-input and eight-input diagnostic results are shown in Tables 6 and 7, respectively. Some observations can be found with these results:

1 success rates decrease with the increase of data scatter;
2 success rates of eight-input ANN are only marginally higher than those of the four-input ANN;
3 FP14, LPC, 25BL, and 29BH faults are more difficult to isolate than others;

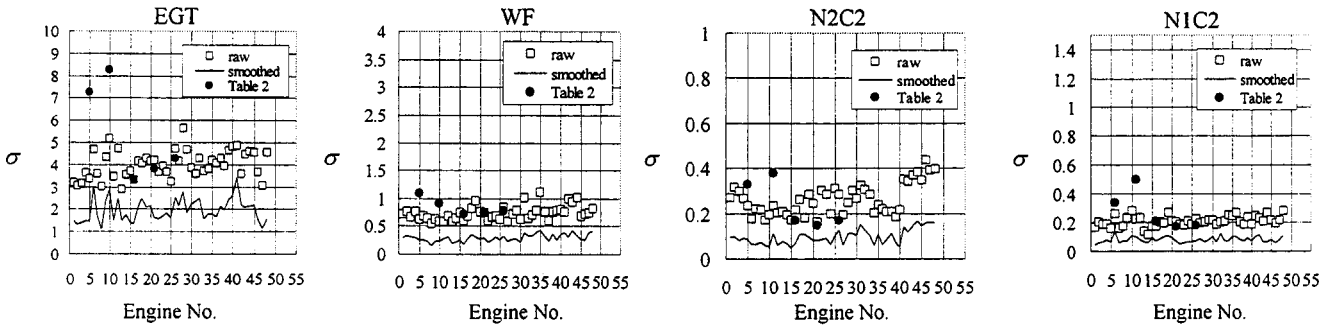


Fig. 1 Comparison of engine scatters

Table 4 Averaged standard deviations of AMECO data

Parameters	T49C2	WF	N2C2	N1C2
Raw Data σ	3.97℃	0.71%	0.24%	0.20%
Smoothed Data σ	1.95℃	0.28%	0.09%	0.08%

4 to meet the requirement of a minimum overall success rate of 90 percent, the parameter K cannot be greater than 3.0 and 3.5, respectively, for the four-input and eight-input diagnoses.

Blind Test Results. A blind test problem set consisting of 150 faults was generated by P&W to evaluate the present ANN-based diagnostic method. BP neural networks with four-input and eight-input nodes, respectively, and 15 output nodes (the first 15 faults in Table 1) were trained. At most three predicted root causes can be proposed for each diagnosis provided these causes are fired up at the output nodes. The BP net is trained by 300 samples and the outcome is shown in Table 8. Both four- and eight-input diagnoses achieve high success rates. A further examination of the failed cases indicates that:

1 most of the prediction errors were between HPT and TCC, and between the bleed systems. To the knowledge of authors, this type of isolation between similar faults has not been satisfactorily resolved by other methods either.

2 it is interesting to see that the success rates of eight-input diagnosis were lower than those of the four-input. The reason, perhaps, is attributed to either that the number of 150 test cases is not statistically sufficient to draw a general conclusion, or that the data scatters of some sensors, for instance, T25C2 and T3C2, are so high that some of the fault features are overly distorted.

Table 5 Test results for BP trained using normal data scatter

No.	Faults	Success Rates K=2.0 (σ =Case 5)	No.	Faults	Success Rates K=2.0 (σ =Case 5)
1	HPC	91%	12	+EGT	75%
2	HPT/TCC	98%	13	+WF	97%
3	25 BS	96%	14	25BL	82%
4	FP14	88%	15	29BH	92%
5	FP8	90%	16	29B2	100%
6	LPC	65%	17	SVM	83%
7	LPT	99%	18	+P49	99%
8	FAN	95%	19	+P2	100%
9	29B1	100%	20	+TAT	99%
10	+N2	74%	21	+Ma/+ALT	100%
11	+N1	99%			
Total Samples Tested: 2200 Averaged Success Rate: 91.0%					

Table 6 Diagnostic success rates of four-input ANN

No.	Type	K=1.0	K=1.5	K=2.0	K=2.5	K=3.0	K=3.5
1	HPC	100%	100%	98%	92%	82%	80%
2	HPT	100%	100%	100%	100%	96%	98%
3	25BS	100%	100%	100%	100%	100%	98%
4	FP14	100%	94%	88%	82%	74%	66%
5	FP8	100%	96%	96%	96%	92%	96%
6	LPC	96%	96%	86%	84%	74%	72%
7	LPT	100%	100%	100%	100%	100%	100%
8	FAN	100%	98%	100%	98%	100%	98%
9	29B1	100%	100%	100%	100%	98%	94%
10	TCC	100%	100%	100%	98%	96%	92%
11	25BL	98%	72%	76%	70%	60%	56%
12	29BH	100%	94%	84%	80%	64%	50%
13	29B2	100%	100%	100%	100%	100%	100%
14	P49	100%	100%	100%	100%	100%	100%
15	SVM	100%	100%	100%	100%	100%	100%
Overall		99.6%	96.6%	95%	93%	89%	87%

Table 7 Diagnostic success rates of eight-input ANN

No.	Type	K=1.0	K=1.5	K=2.0	K=2.5	K=3.0	K=3.5
1	HPC	100%	100%	100%	100%	98%	98%
2	HPT	100%	100%	100%	100%	100%	100%
3	25BS	100%	100%	100%	100%	100%	100%
4	FP14	100%	100%	96%	92%	92%	64%
5	FP8	100%	100%	94%	96%	92%	84%
6	LPC	100%	98%	94%	86%	92%	86%
7	LPT	100%	100%	100%	100%	100%	100%
8	FAN	100%	98%	98%	96%	98%	96%
9	29B1	100%	100%	100%	100%	100%	100%
10	TCC	100%	100%	100%	100%	100%	92%
11	25BL	100%	100%	82%	84%	62%	62%
12	29BH	100%	98%	92%	88%	84%	78%
13	29B2	100%	100%	100%	100%	100%	100%
14	P49	100%	100%	100%	100%	100%	98%
15	SVM	100%	100%	100%	100%	100%	100%
Overall		100%	96.6%	97%	96%	94%	91%

Table 8 Blind test results

	4-input/15 Root Causes	8-input/15 Root Causes
Accuracy of First Pick	91%	88%
Accuracy of First Two Picks	94%	94%
Accuracy of Three Picks	96.3%	94.7%

Sensor Validation Neural Network

Autoassociative ANN. It has been clearly demonstrated in the previous four- and eight-input fault diagnostic studies that the success rate depends virtually on the quality of the sensor data. For an ideal case of using noise-free training and testing data, the success rate can be 100 percent. Therefore the design of a preprocessor that can effectively reduce the input noise level is of paramount importance.

Autoassociative neural network attracts much attention in the recent real-time controller design. It has the advantage of fast computing and can perform sensor failure detection, identification, and accommodation (SFDIA) simultaneously. Of its functions for noise filtering, missing sensor replacement, and gross error detection and identification, we only examine the noise-filtering ability and reserve the other capabilities for future investigations. In the practical application of engine condition monitoring, a sensor could be biased or failed without appreciation. Invalid sensory information is often the source of false alarm as data are inserted without examination into the diagnostic system.

The autoassociative ANN used is a feedforward network, as illustrated in Fig. 2. A major distinctive feature of the autoassociative network is that it contains a bottleneck layer. The number of bottleneck nodes should be lesser than that of the input nodes, but larger than the degree of freedom of the system being diagnosed. The input and output layers have identical parameters. Theoretically, autoassociative ANN forms a unitary mapping which maps the input parameters onto themselves. All the noise-filtering process is undertaken in the mapping part, namely from the input to the bottleneck layer. Once the intrinsic states are retrieved from the mapping process, a demapping part is followed by decoding the bottleneck nodes back to the output nodes. In our autoassociative architecture design we use BP algorithm with sigmoidal nonlinearity assigned to each hidden nodes including the bottleneck nodes.

Eight-input AANN Results and Discussions. The autoassociative network selected for discussion has a structure of 8-9-5-9-8 in which 5 denotes the number of bottleneck nodes. In gen-

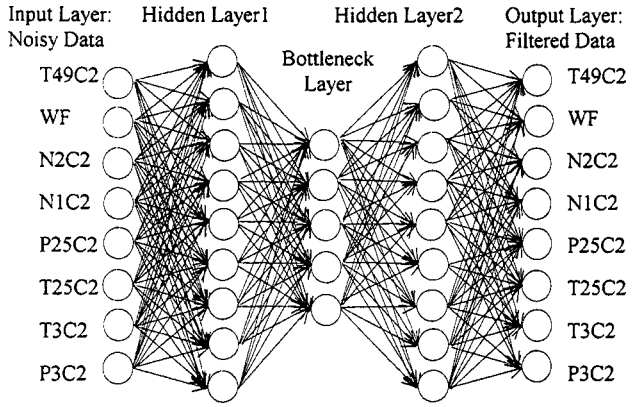


Fig. 2 Autoassociative neural network architecture (8-9-5-9-8) used for eight-input data filtering

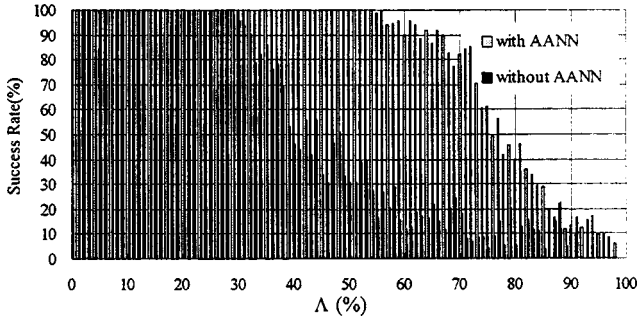


Fig. 3 Success rate vs noise-to-signal ratio for input data with/without autoassociative neural network (AANN) filtering

eral, the convergence rate of autoassociative ANN is slow, and at present there is no conclusive acceleration method that is found effective for improving the present AANN BP algorithm.

In the training of the autoassociative ANN, noise-free data are used at the output nodes while various kinds of noisy input data are inserted from the input nodes. It could be argued that in the real-world application, the noise-free data can never be available. Nonetheless, the optimal condition of using noise-free data for the output nodes is the limit an autoassociative network can do for noise filtering. It could be anticipated that using the same noisy data at both input and output nodes will result in a less effective noise filtering autoassociative network. The remaining question is about how low the noise level should be to make an acceptable noise-filtering AANN.

Although the effectiveness of noise filtering can be evaluated by comparing the data prior to and after filtering, however, often the noise-filtering effect is not uniformly valid for every sensor data. For the present AANN architecture, it is hard to filter out noise contained in the N2C2 data while others are almost completely recovered noise free. In order to give a fair evaluation of the autoassociative filter, we use the success rate of the diagnosed results as the measure of filter effectiveness. In other words, the autoassociative net is connected onto the BP diagnostic network and we use the diagnosed output results of the integrated nets as the basis for comparison. The BP diagnostic network is trained using scatter-free data. Therefore the reduction of noise level is mainly accounted in the autoassociative filter.

Illustrated in Fig. 3 are the probability density function type exemplification of the noise-filtering effect. The abscissa variable Λ measures the noise-to-signal ratio and is defined by

$$\Lambda = \frac{1}{M}\sum_{i=1}^{M} \Lambda_i \tag{2a}$$

$$\Lambda_i = \frac{N_i}{N_i + S_i} \tag{2b}$$

$$S_i = |\overline{X_i}|, \quad N_i = |X_i - \overline{X_i}| \tag{2c}$$

in which X_i is the ith sensor data and $\overline{X_i}$ the ith scatter-free data. The integer M represents the number of sensors and presently we choose $M=8$ for the eight-input network. Note that the parameter Λ can be taken as a noise-to-signal ratio, $0 \leqslant \Lambda \leqslant 1$, and the larger the noise the bigger the Λ value. The ordinate variable is the percentage of successful fault diagnosis for samples having noise levels in the range between Λ and $\Lambda+\Delta\Lambda$, and in Fig. 3 $\Delta\Lambda = 0.05$.

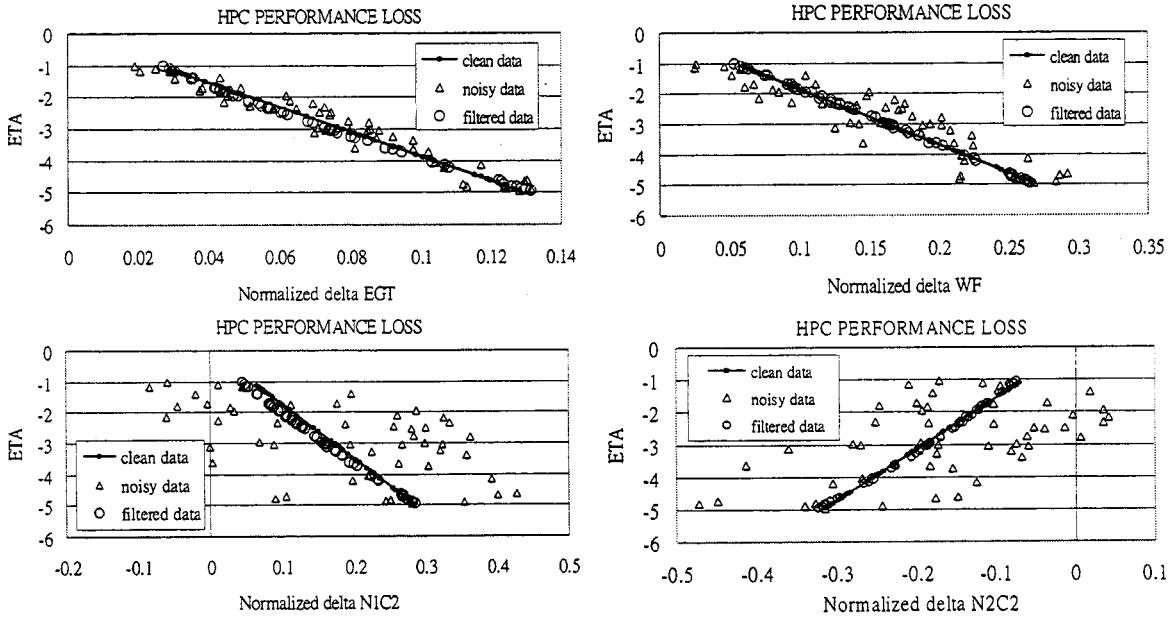


Fig. 4 Noise-filtering capability of AANN

Sensor data are first generated according to Eq. (2), in which K ranges from 0 to 600. The generated data are rearranged, from small to large numbers, into a queue using the definition Λ and then diagnosed and converted into Fig. 3.

Figure 3 shows the results obtained with and without autoassociative filtering. For diagnosis without autoassociative filtering, the success rate above 90 percent corresponds to $\Lambda \leqslant 33$ percent. The tolerance of the diagnostic network can be doubled, $\Lambda \leqslant 68$ percent, when the autoassociative filter is implemented. This is seen from the much elevated distribution shown in Fig. 4. It is believed that the present autoassociative network can be further expanded and elaborated to include in the functions of missing data replacement and sensor bias correction, making the present neural network approach more robust for the future real-world applications.

AANN as Trend Detector. A timely detection of trend changes is important for engine condition monitoring. However, of equal importance is the avoidance of false alerts. AANN is a data validity preprocessor trained by fault samples with various levels of trend changes. Therefore AANN can serve as a trend detector while analyzing the input data validity. DePold and Gass (1999) [18] proposed to detect trend using ANN trained by three sets of filtered data. It is claimed that 66 percent of saving in detection time response can be achieved. Noise filtering and trend detection can be done simultaneously using the present AANN and, in principle, the trend can be detected with no time delay, as explained below.

Figure 4 shows the noise-filtering capability of the presently trained AANN. The training samples used cover a range of HPC efficiency delta from -1 percent to -5 percent. It is shown that the noise can be effectively filtered to result in a filtered data trend which is much closer to the original noise-free straight line behavior. Also shown in Fig. 5 is a time series representation of this AANN trend detection procedure. A 2 percent trend change of HPC performance loss was assumed, and on the clean data a data scatter corresponding to $K=2.0$ is superimposed. As conventional ten-point moving average or its variant of exponential memory retention (DePold and Gass, 1999 [18]) is applied to these noisy raw data, it takes 10–12 points for trend change to be revealed. However, contrary to these conventional smoothing methods, AANN can detect this trend immediately without time delay.

To further examine the data validity capability of AANN, several unreasonably large deviants or ''wild'' points were purposely implanted in the time series. It can be seen in Fig. 6 that no false alarm was predicted by the present AANN filter as well.

From the above observation, it is proposed that AANN can be used as a good candidate for data preprocessing. Not only sensor error can be detected and identified, the subsequent fault isolation can be improved to a higher rate of success, most importantly, the trend can be detected instantaneously which makes the neural network-based condition monitoring a more valuable tool.

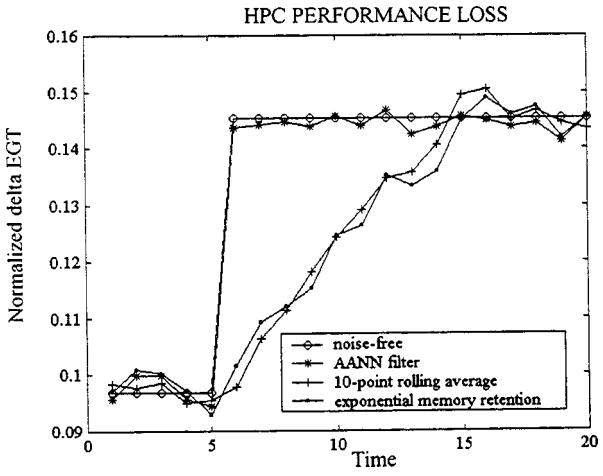


Fig. 5 Comparison of data smoothing methods in trend detection

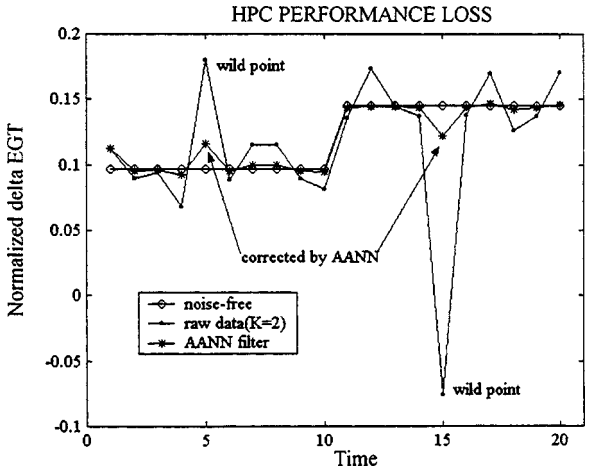


Fig. 6 ''Wild'' points data correction by AANN filter

Concluding Remarks

The collaborative research on developing artificial neural network fault diagnostic method has been carried out with initial success. Both four-input and eight-input networks were constructed using fault samples generated by the P&W supplied influence coefficient matrix and data scatters. A proper noise model which reflects the actual parameter scatter characteristics was obtained after a study of the real engine data. The objective of the present investigation is to evaluate the ability of the ANN-based method in diagnosing the selected 15 faults. A minimum 90 percent success rate is set as the criterion for this diagnostic method to be considered acceptable. The conclusions drawn from this collaboration are as follows:

1 Both four-input and eight-input BP neural networks meet the requirement of a minimum 90 percent success rates.

2 The unidentifiability problem of the model-based linear system approach is circumvented. A systematic diagnostic process can be done with the limited four-input system, and the results are almost as good as those of the expanded eight-input system.

3 An effective preprocessing method for validating the raw data is of paramount importance. Preliminary study of autoassociative neural network has been performed for sensor data validation. It is an effective noise filtering and instrumentation fault isolation preprocessor. The tolerance of the present diagnostic BP network to data scatter can be significantly enhanced as autoassociative filter is used.

4 The autoassociative neural network can serve as an effective trend detector which in principle can detect trend changes without any time delay.

5 Engine performance might migrate when service time or cycles accumulate. Method of adapting the parameters of an engine simulation code so as to let the influence coefficient matrix reflect and keep up with the actual engine operational condition should be considered. This adaptive influence coefficient matrix will be indispensable for the practical use of the present methodology.

6 Further examination of the presently developed neural network method with actual fault cases is necessary. The high success rate ought to be further verified using actual engine data.

Acknowledgments

The present work is supported by a joint collaboration contract sponsored by P&W. The authors are deeply indebted to Hans Depold, Ranjan Ganguli, and Alan Volponi of P&W for many insightful discussions and suggestions. The continuing encouragement of Dr. Bob Ni, P&W senior fellow, which made this collaboration materialized is mostly appreciated.

References

[1] Urban, L. A., 1972, ''Gas Path Analysis Applied to Turbine Engine Condition Monitoring,'' AIAA/SAE Paper 72-1082.
[2] Urban, L. A., and Volponi, A. J., 1992, ''Mathematical Methods of Relative Engine Performance Diagnostics,'' *SAE 1992 Transactions*, Vol. 101, *Journal of Aerospace*, Technical Paper 922048.
[3] Doel, D. L., 1994, ''TEMPER-A Gas Path Analysis Tool for Commercial Jet Engines,'' ASME J. Eng. Gas Turbines Power, **116**, No. 1, pp. 82–89.
[4] Doel, D. L., 1994, ''An Assessment of Weighted Least-Squares-Based Gas Path Analysis,'' ASME J. Eng. Gas Turbines Power, **116**, No. 2, pp. 366–373.
[5] Zedda, M., and Singh, R., 1998, ''Fault Diagnosis of a Turbine Engine Using Neural Networks: a Quantitative Approach,'' Paper No. AIAA-98-3602.
[6] Whitehead, B., Kiech, E., and Ali, M., 1990, ''Rocket Engine Diagnostics Using Neural Networks,'' Paper No. AIAA-90-1892.
[7] Eustace, R., and Merrington, G., 1995, ''Fault Diagnosis of Fleet Engines Using Neural Networks,'' ISABE 95-7085, pp. 926–936.
[8] Roemer, M., and Atkinson, B., 1997, ''Real-Time Health Monitoring nd Diagnostics for Gas Turbine Engines,'' ASME Paper No. 97-GT-30.
[9] Cifald, M. L., and Chokani, N., 1998, ''Engine Monitoring Using Neural Networks.'' Paper No. AIAA-98-3548.
[10] Werbos, P. J., 1974, ''Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences,'' Ph.D thesis, Harvard University, Cambridge, MA.
[11] Haykin, S., 1994, *Neural Networks, A Comprehensive Foundation*, Macmillan, New York.
[12] Romer, M., 1998, ''Testing of a Real-Time Health Monitoring and Diagnostics System for Gas Turbine Engines,'' Paper No. AIAA-98-3603.
[13] Kramer, M. A., 1991, ''Non-linear Principal Component Analysis Using Autoassociative Networks,'' AIChE J., **37**, No. 2, pp. 233–243.
[14] Kramer, M. A., 1992, ''Autoassociative Neural Networks,'' Comput. Chem. Eng., **16**, No. 4, pp. 313–328.
[15] Napolitano, M., Windon, D., Casanova, J., and Innocenti, M., 1996, ''A Comparison between Kalman Filer and Neural Network Approaches for Sensor Validation,'' Paper No. AIAA-96-3894.
[16] Mattern, D. L., Jaw, L. C., Guo, T. H., Graham, R., and McCoy, W., 1997, ''Simulation of an Engine Sensor Validation Scheme Using an Autoassociative Neural Networks,'' Paper No. AIAA-97-2902.
[17] Mattern, D. L., Jaw, L. C., Guo, T. H., Graham, R., and McCoy, W., 1998, ''Using Neural Networks for Sensor Validation,'' Paper No. AIAA-98-3547.
[18] 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**, No. 4, pp. 607–612.

Nonlinear Model Predictive Control Experiments on a Laboratory Gas Turbine Installation

H. A. van Essen
e-mail: h.a.v.essen@tue.nl

H. C. de Lange

Department of Mechanical Engineering,
Eindhoven University of Technology,
PO Box 513,
5600 MB Eindhoven, The Netherlands

Results on the feasibility and benefits of model based predictive control applied to a gas turbine are presented. For a laboratory gas turbine installation, the required dynamic simulation model and the real-time (nonlinear) model predictive control (MPC) implementation are discussed. Results on both model validation and control performance are presented. We applied a nonlinear MPC configuration to control the laboratory gas turbine installation and succeeded in a real-time implementation. Although the available computation time for prediction and optimization of the model limits the sample time, the advantages of MPC, i.e. constraint handling, and anticipation to future (set-point) changes are fully reached, and the control performance is good. Special attention is paid to the performance of the applied filter that compensates for inevitable mismatches between model and process measurements. In general, the opportunities of model based control of turbomachinery are promising. [DOI: 10.1115/1.1359478]

Introduction

Model predictive control is a control strategy that uses a model of the installation to predict the response over a future interval, called the prediction horizon. Future control inputs are determined through minimizing a customized criterion, *e.g.* deviation from a desired set-point, over (a part of) this future interval, the control horizon. Of the computed optimal control moves, only the values for the first sample are actually implemented and the algorithm repeats the same procedure over the next sample. The main benefit of MPC is its constraint handling capacity: unlike most other control strategies, constraints on inputs and outputs can be incorporated into the MPC optimization. Another benefit of MPC is its ability to anticipate to future events as soon as they enter the prediction horizon. Finally, MPC is an essentially multivariable control strategy, implying that control loops do not need to be decoupled, because all interactions between multiple inputs and outputs are accounted for by the model.

For gas turbine control, this implies that physical operation limits, *i.e.* compressor surge, maximum allowable expander inlet temperature or rotational speed, as well as limits on actuator actions can be included into the optimization leading to optimal control of changes in operating point even when constraints are (temporarily) reached.

In this study, we consider a small (400 kW thermal input), custom built, laboratory gas turbine setup. This installation serves for experimental validation of simulation models and to test real-time control implementations. The laboratory installation offers the opportunity to monitor and to influence the dynamic operation of a gas turbine. Industrial configurations are not available yet for these experiments. Indeed, the laboratory setup is a fair representation of a gas turbine installation. The nonlinearities, dynamics, and constraints as well as the properties of the composing components correspond to industrial systems. Relevant time scales, however, differ as the components have smaller dimensions.

The application of MPC to control this laboratory gas turbine has been introduced in Essen [1] and Vroemen [2]. Vroemen [2] introduced the standard linear MPC concept, worked out a nonlinear method based on successive linearization, and presented preliminary simulation results. Successive linearization implies that every sampling period an updated linear model is derived from an underlying nonlinear model. This linear model is used for optimization over the control horizon. This approach has been shown to be a useful and powerful extension to linear MPC.

This paper presents experimental results obtained with the nonlinear MPC implementation on the laboratory gas turbine. In subsequent sections, first the laboratory installation is briefly introduced, after which the validation of the model (parameters) is discussed, the specifics of the MPC implementation are treated, and some typical results are addressed. The paper is concluded by some recommendations for future research in model based control of turbomachinery.

The Laboratory Gas Turbine

The laboratory gas turbine has been built from an industrial turbocharger and a custom made combustion chamber. The turbocharger comprises a single stage radial compressor and a single stage axial expander. Figure 1 presents a schematic view, showing the compressor, a blow-off valve, a check-valve, a buffer tank, a throttle valve, the combustion chamber, and finally the expander. A compressed air facility is employed in the start-up procedure, but also serves to disturb to the nominal operation point to excitate the control system.

Three electrically powered valves provide opportunities to control the operation point of the gas turbine. The blow-off valve, however, is only used during startup. Due to its large capacity, it is not suited as a control input. The throttle valve influences the pressure ratio over compressor delivery and expander inlet. As no external load is connected to the shaft, the throttle valve simulates the load condition of the installation. The throttle valve, therefore, can either be used as an input to control the operation point OR as a system parameter. The fuel valve controls the power supplied to the gas turbine and is the most important control input. Due to the mechanical transmissions in the valves, the rate of change is limited to full stroke in 40 s for the fuel valve and even 120 s for the throttle valve.

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Munich, Germany, May 8–11, 2000; Paper 00-GT-040. Manuscript received by IGTI Feb. 2000; final revision received by ASME Headquarters Jan. 2001. Associate Editor: M. Magnolet.

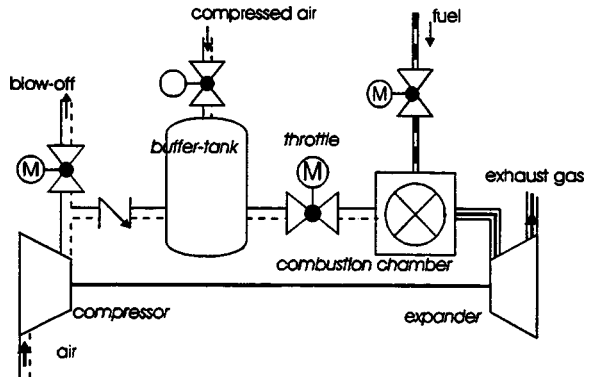


Fig. 1 Schematic view of the gas turbine installation

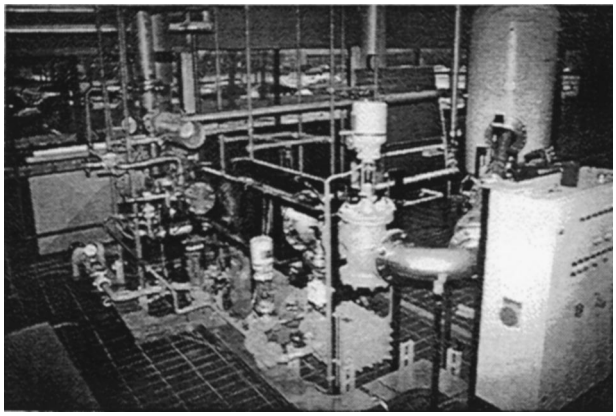


Fig. 2 Photograph of the gas turbine installation

A LabVIEW based data-acquisition system has been realized. This system provides opportunities to measure, monitor and store the available measurement signals, including pressures, temperatures, valve positions, rotational speed, and mass flow through a built-in orifice. The LabVIEW program is also able to implement new valve positions during automatic control.

Figure 2 presents a photograph of the installation showing the large buffer-tank (height over 2 meters) on the back and the combustion chamber and the throttle valve, above the fuel supply line, in front. For detailed information, the reader is referred to Essen [2].

Dynamic Simulation Model

A physical simulation model has been developed. This model is preliminarily introduced in Vroemen [3] and described in detail in Essen [1]. The model has a one-dimensional, generic, modular, component-wise structure. Components are compressor, expander, control valves, combustion chamber, flow-restrictions, and piping. The physical basis and the modular approach ensure the applicability of the model for a large class of compressor/expander systems.

Components are modeled by their characteristics, *i.e.* a set of (nonlinear) steady algebraic equations determining the mass flow through the component as a function of pressure, temperature, rotational speed, etc. In this way the performance characteristics are used as static momentum balances. The dynamic behavior of the components is modeled by volumes, positioned in between two components. Inside each volume the conservation equations *for mass*, *momentum*, and *energy* are solved, *i.e.* two coupled ordinary differential equations in pressure and temperature.

The model configuration for the laboratory gas turbine in Fig. 3 shows three volumes representing the effective volumes of the

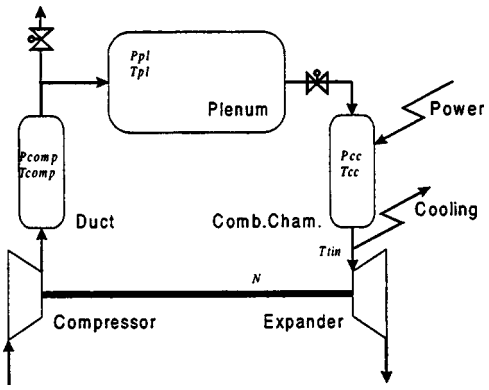


Fig. 3 Model configuration with coupling volumes between successive components

compressor duct, the buffer-tank plus piping (plenum), and the combustion chamber. The describing equations for each volume are

$$\frac{dT}{dt}=\frac{RT}{pV}\left[\gamma\left(\dot{m}_{\text{in}}\frac{C_{p,\text{in}}T_{\text{in}}}{C_p}-\dot{m}_{\text{out}}T+\frac{Q}{C_p}\right)-T(\dot{m}_{\text{in}}-\dot{m}_{\text{out}})\right]$$

$$\frac{dp}{dt}=\frac{\gamma R}{V}\left(\dot{m}_{\text{in}}\frac{C_{p,\text{in}}T_{\text{in}}}{C_p}-\dot{m}_{\text{out}}T+\frac{Q}{C_p}\right)$$

In which the term Q represents the added combustion power or the extracted expander cooling power, respectively. A power balance on the rotational speed of the gas turbine shaft gives the seventh state equation.

System simulation based the performance maps of the components extended by dynamic coupling equations is widely accepted in the literature and presented, among others, by Schobeiri [4] and Garrard [5] with regard to aero engines and Botros [6,7] with regard to compressor stations and piping. Apart from the characteristics, only little detailed (geometric) system information is required.

For control purposes, fast simulation is required and the model configuration is kept as simple as possible. Therefore, piping and buffer tank have been lumped into a single plenum volume. Due to its relative large size (>2 m^3), the cut-off frequency (determined by the residence time of the largest volume) of the model is estimated at 0.2 Hz. When transients at a faster time scale need to be controlled, the model may fail. In that case, the plenum should be split into multiple volumes in series, combined with an instationary momentum balance, Essen [1], allowing compressible flow at the expense of a considerably larger simulation time. As the required time scale for controlling the operation point (approximately 20 s) corresponds to this limit, our model of the laboratory gas turbine is suitable for application in a model based controller.

Parameter Validation

Due to the physical nature of the model, a large number of parameters is involved. Two types of parameters are distinguished. The first type influences the stationary operating points of the installation. These parameters include pressure drop factors and efficiencies. *Pressure drop factors* determine the mass flow through components (often hidden in the component characteristics). *Efficiencies* include the (polytropic) efficiencies of compressor and expander, and the combustion efficiency. These steady state parameters have been validated by extensive experiments in a systematic procedure based on a mass and power balance.

During the validation of the steady-state parameters we had to modify the compressor performance map (available from the original manufacturer) to have the map agree with the experimental results. Reasons for this mismatch are not fully understood but

may be in the interaction between components and in the methods by which the maps of separate components are determined.

Due to the lumped influence of the expander cooling system, the neglected temperature dependency of the heat capacity, and the inherent dynamic character of the combustion process, it appeared difficult to determine the efficiencies of expansion and combustion. In the present model, the combustion efficiency is a static function of the fuel-air ratio. This leads, however, to significant (dynamic) model mismatches in the expander inlet temperature when suddenly more fuel is injected.

The second type of parameters influences the transient operation of the installation. These parameters include the size of the effective volumes, the inertia of the gas turbine shaft, and the move rates of control valves. The transient behavior did not introduce any problems; all the inertia parameters were easy to select from straightforward (physical) start values.

With validated parameters, the performance of the model is good. Both steady state and transient simulations agree well with experimental results.

Nonlinear MPC Implementation

For the gas turbine setup the emphasis is on real-time implementation of MPC and we defined rather straightforward control objectives as set-point and trajectory control, combined with constraints on inputs and outputs. For our MPC configuration we use Primacs, a package for real-time model based control that is being developed by TNO-TPD (Delft, The Netherlands). Although Primacs was originally designed for linear MPC, the modular structure allowed us to implement the *successive linearisation* approach. The connection to the LabVIEW environment has been established by DDE over an Ethernet connection. We succeeded in a real-time implementation of MPC on the laboratory installation. The sample interval was limited to 1.2 s on a ordinary Pentium 200 PC.

Because of inevitable model mismatches and measurement noise or errors, a *filter* is required to correct the prediction of the internal model towards the actual process measurements. In our implementation we tested a first order integrating (output disturbance) filter. In this filter, a number of extra states (d) track the offset between process measurements and model predictions. The filter equations are given by

$$\begin{bmatrix} x(k+1) \\ d(k+1) \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} x(k) \\ d(k) \end{bmatrix} + \begin{bmatrix} 0 \\ K_f \end{bmatrix} [y_m(k) - y_p(k)] + \begin{bmatrix} B \\ 0 \end{bmatrix} u(k)$$

$$y_m(k+1) = [C \quad I] \begin{bmatrix} x(k+1) \\ d(k+1) \end{bmatrix} + Du(k)$$

in which y_p is a process measurement and y_m is the corresponding model prediction. Note that only the outputs and not the states are updated. A basic extension to this scheme to avoid linearization errors in the filter correction is a one-sample-ahead nonlinear prediction for y_m.

This first order filter performs well as long as all controlled outputs can be measured (or accurately reconstructed by other measurements). Otherwise, the filter cannot correct the model predictions accurately and the controller performances degrades. In our configuration, a problem arises with the reconstruction of the mass flow through the compressor. As it is impossible to measure it directly it should be reconstructed from the compressor characteristic. This reconstruction, however, is not only very bad conditioned by almost flat curves in the compressor map near the surge line, but it may also be corrupted by deviations between measured and modeled *states* (rotational speed, pressure ratio). Although we did not attempt this, an augmented Kalman filter, that updates the model *states* instead of the *outputs*, could possibly improve this problem. It is, however, not easy to implement in a nonlinear setting [1] and requires sufficient measured outputs to feedback.

Results

In this section some typical experimental results are presented and discussed with emphasis on constraint handling and robust performance. Constraint handling appears to be a realistic advantage of MPC for control of turbomachinery. Most ''conventional'' controllers cannot deal with constraints at all, especially not when constraints (like surge and expander temperature) are not on *controlled* outputs. For the laboratory installation, surge and expander temperature constraints are reached frequently. Also (move) constraints on the actuator inputs strongly influence the controlled operation. Figure 4 displays the operating area of the gas turbine installation in the compressor characteristic, indicating the constraints.

Note that the ''surge-ratio'' constraint is a safety margin to the real surge line defined by the ratio of actual compressor mass flow and the corresponding mass flow at the surge line for the same pressure ratio.

Sine Responses. This result illustrates the capabilities of the MPC controller to track a periodic reference trajectory. A sinusoidal trajectory for the rotational speed (period 75 s, amplitude 80 rev/s, mean 438 rev/s) is specified while the throttle valve serves as a (known) load-disturbance on the installation (sine with half frequency: period 150 s, amplitude 0.3, mean 0.7). The fuel valve is now the only input. In Fig. 5*a* both sine functions are presented. Figure 5*b* shows the tracking result over a time interval of one full period of the disturbance in a steady state response, that is, after the switch-on phenomena vanished.

Indeed, Fig. 5*b* shows that the controller is able to follow the trajectory on the rotational speed, except for the two rising flanks. There, the temperature constraint is reached, as can be seen in Fig. 6*a*, which displays the expander temperature over the same time interval. Moreover, a difference can be seen between the first

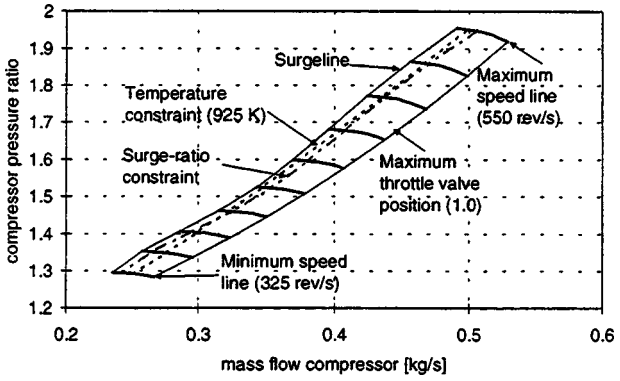


Fig. 4 Simulated operation area, limited by constraints, projected onto the compressor map

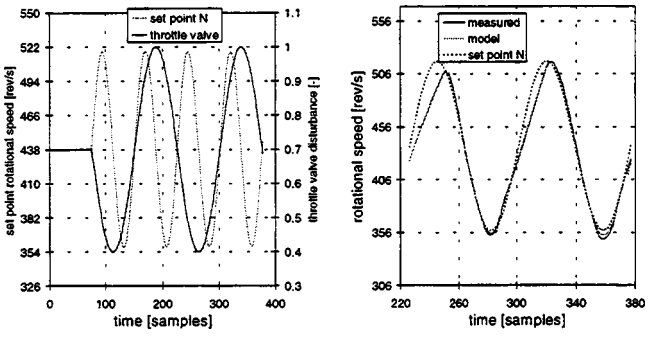


Fig. 5 (*a*) Sinusoidal trajectories for the rotational speed and throttle valve disturbance. (*b*) Measured and predicted responses of rotational speed.

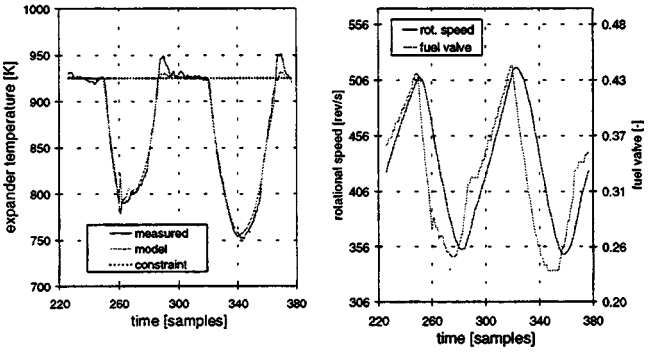


Fig. 6 (*a*) Expander temperature during sine response experiment. (*b*) Comparison between control output (rotational speed) and input (fuel valve).

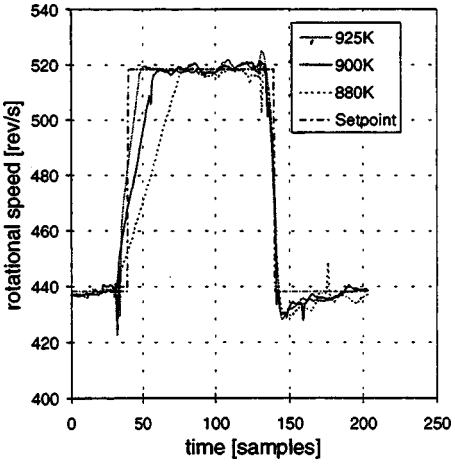


Fig. 8 Rotational speed for different settings of the temperature constraint

rising and the second rising flank. This difference is caused by the opposite influence of the throttle valve (*i.e.*, gas turbine load) disturbance in these two situations.

Figure 6*a* shows that the controller indeed manages to keep the temperature within the constraint. Clearly, however, a temporary constraint violation of the measured expander inlet temperature can be seen. As the controller is able to keep the (filtered) model value within the constraint, this appears to be a filter problem. The controller is simply not aware of the violation because the filter is not fast enough to compensate the mismatch of the expander temperature. Indicated solutions are in a higher filter gain for this output, a faster MPC sampling time, and, of course, in an improved modeling of the (dynamic) combustion efficiency.

Figure 6*b* compares the actual rotational speed with the fuel valve input. This illustrates the anticipative behavior of the controller. Already 12 samples (one prediction horizon) before a change in sign of the derivative of the rotational speed, the fuel valve changes direction. Anticipation will be discussed in a separate subsection.

Surge or Temperature Constraint. Close to the surge line, depending on the rotational speed level, the simulated operating area in Fig. 4 is either restricted by the surge-ratio constraint or the temperature constraint. This has been verified in two experiments, at two different constant rotational speeds. It is examined whether the controller ''stops'' at the surge constraint or at the temperature constraint when the mass flow through the expander is forced to decrease. The experiments involve a constant set-point on the rotational speed and a decreasing set-point on the mass flow. As controlled inputs, both the fuel and the throttle valve are used. The desired operating points are located across the surge line. The corresponding constraints are indicated in Figs. 7*a* and *b*, showing that at the high rotational speed (438 rev/s), the controller stops at the surge ratio constraint of 0.97, while at the low rotational speed (350 rev/s), the controller stops at the temperature constraint of 925 K. These experimental results compare to the expected results from the simulation and show that the MPC controller is able to maintain constraints on outputs that are not controlled.

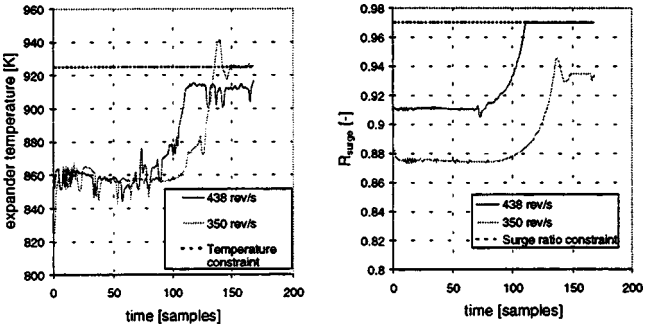


Fig. 7 (*a*) and (*b*) Expander inlet temperature and surge ratio parameter for two experiments with constant rotational speed and decreasing mass flow

Temperature Constraint Level. Results on the influence of the level of the temperature constraint on the controlled variable rotational speed are presented in Fig. 8. For three different values of the temperature constraint (880, 900, and 925 K), the same steps are put on the set-points of the rotational speed and the expander mass flow (not shown). Lowering the maximum allowable expander temperature results in a slower system response. Obviously, the allowed increase of the fuel flow per sample has to be reduced. As expected, the temperature constraint does not influence the response to a lower set-point. In fact, these responses are approximately equal. Figure 9 indicates that the controller manages to keep the constraints, although they are temporarily violated as discussed before.

Disturbances. A rigorous way to apply a disturbance to the system is the injection of compressed air into the buffer tank by (manually) opening the valve (Fig. 1). In the experiment presented in Fig. 10, the system response for constant rotational speed and expander pressure is examined. The controller cannot anticipate to this unmodeled disturbance and all deviations should be compen-

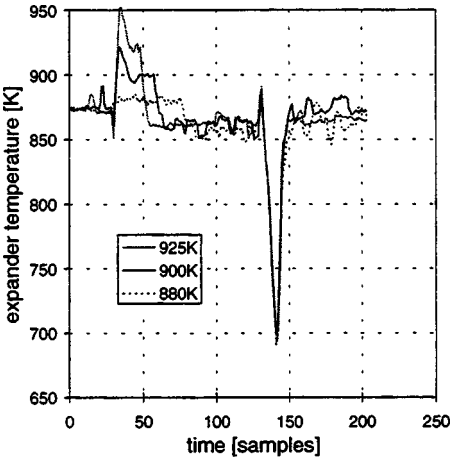


Fig. 9 Expander inlet temperature

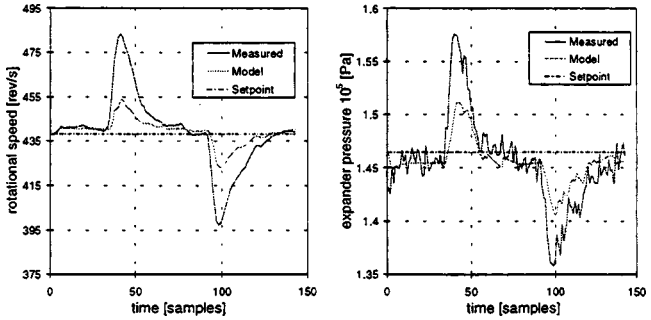


Fig. 10 (*a*) and (*b*) Predicted and measured rotational speed and expander pressure during compressed air disturbance

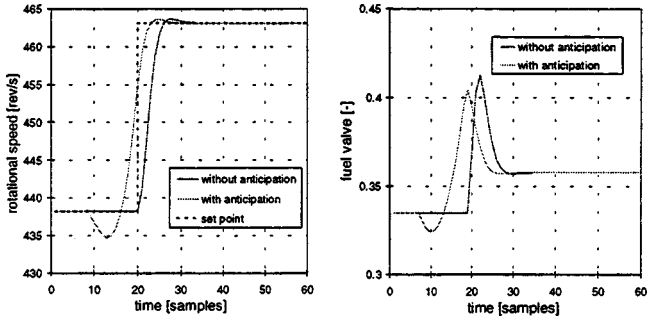


Fig. 12 (*a*) and (*b*) Simulated response of rotational speed with and without anticipation

sated by the filter. At sample 35 the compressed air valve is opened, while at sample 90 the valve is closed again. The injected mass flow of compressed air is approximately 0.05 kg/s (or 12 percent of the expander mass flow). Although large filter corrections must be applied, the controller manages to reach the original set-points of both speed and pressure again. The system is therefore robust for large disturbances and the filter performs well at this time scale.

Transient Response. In the simulation shown in Fig. 11, the transient behavior during a set-point change is indicated in the (compressor) characteristic. A new set-point has been specified at a lower rotational speed, while the fuel valve is the only input. It can be seen that the surge constraint is nearly reached by the compressor mass flow. Note the difference between compressor and expander mass flows due to depressurizing the buffer tank. For a gas turbine like the laboratory installation, there is a potential danger of surge during fast ''downwards'' transients. A typical example is emergency shutdown. A model based controller is able to anticipate to this problem.

Anticipation. Anticipation is the ability of the controller to react before an actual set-point change is commanded. In this subsection we present some results on anticipative behavior of model predictive control. In Fig. 12*a*, the simulation results of a system response to a set-point change in rotational speed with and without anticipation are compared. The fuel valve, Fig. 12*b*, is the only input and no constraints on the output nor on the temperature are applied. Without anticipation, the controller starts responding as soon as the set-point change is activated. Remarkable is the ''inverse controller response'' that the system displays in the case with anticipation. First the fuel valve is closed and the rotational speed lowers, then the fuel valve opens again and the rotational speed rises to its new set-point value. The inverse response does not seem to be the optimal response to the set-point change. It appears, for instance, already better when the ''negative'' control moves (in the first few samples) are suppressed.

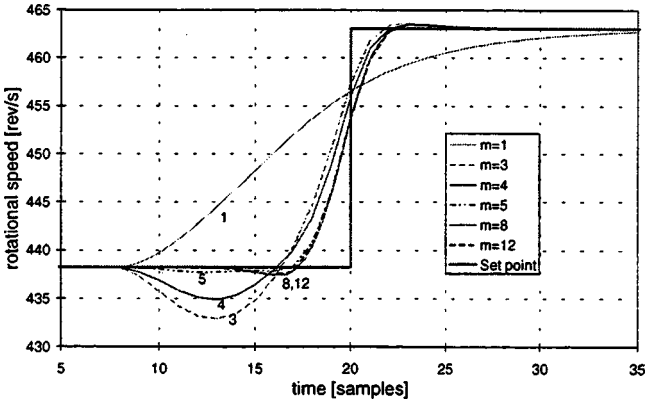


Fig. 13 Response of rotational speed for different settings of the control horizon (*m* samples) for a prediction horizon of 12 samples

This phenomenon of anticipative behavior appears to be strongly related to the relative size of the control horizon with respect to the prediction horizon. The prediction horizon is the number of samples for which the future system is predicted, while the control horizon is the number of samples for which new future inputs are computed (in general the degrees of freedom). Suboptimal solutions are an inherent consequence of differences between these two horizons and cause an ''inverse'' response. These suboptimizations can be improved by increasing the control horizon relative to the prediction horizon. The influence of the size (*m*) of the control horizon relative to the prediction horizon (12 samples) on the simulated performance is investigated and shown in Fig. 13.

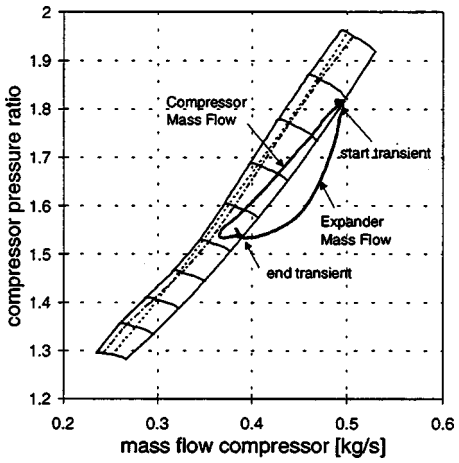


Fig. 11 Simulated transient response displayed in the compressor map

Conclusions

We applied a nonlinear MPC configuration to gas turbine control. We succeeded in a real-time implementation, due to hardware limitations the sample time was limited to 1.2 s. We find that MPC is a challenging, new control strategy for turbomachinery. The performance of the implemented MPC configuration is good. Both steady state set-point variations and transient tracking of reference trajectories are performed well. Constraint limits are observed and the influence of MPC tuning parameters like weighting factors and length of horizons is well understood. The pres-

ence of model mismatches and disturbances is handled correctly by the filter although limited filter performance is responsible for temporary violations of the temperature constraint during fast transients. Anticipation is a powerful control tool, but the undesired phenomenon of inverse response can only be avoided by computationally demanding long control horizons.

To make the advantages and opportunities of model based control available to industrial turbomachinery configurations, more research, both in simulations and in experiments is required.

The following suggestions for future research are made. First, based on advanced hardware, the sample time should be decreased, allowing faster control, this should be combined with the introduction of fast and accurate control valves and new filter implementations. Second, more emphasis is needed on the development of accurate control-oriented models for combustion and especially dynamic combustion efficiency. Third, an extension of the MPC algorithm towards a *mixed integer optimization*, Bemporad [8], should be aimed for. This strategy includes the optimal selection of discrete events like turning machines in or out into the performance criterion of the continuous optimization. An interesting application is reported for load balancing of compressor stations in Essen [1] and Smeulers [9]. Finally, the concept of hierarchical controller implementation, where MPC is not used as the primary controller, but only computes the (optimal) set-points for local controllers, seems promising to maintain the overall optimizing properties of MPC and simultaneously reduce the required computation time.

Acknowledgment

The authors gratefully acknowledge TNO-TPD at Delft, The Netherlands, for their support.

References

[1] van Essen, H. A., 1998, ''Modelling and Model Based Control of Turbomachinery,'' Ph.D. thesis, Eindhoven University of Technology, The Netherlands.
[2] van Essen, H. A., 1995, ''Design of a laboratory gas turbine installation,'' Eindhoven University of Technology, The Netherlands.
[3] 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.
[4] Schobeiri, M. T., Attia, M., and Lippke, C., 1994, ''GETRAN: A generic, modularly structured computer code for simulation of dynamic behavior of areo- and power generation gas turbine engines,'' J. Eng. Gas Turbines Power, **116**, pp. 483–494.
[5] Garrard, D., 1996, ''ATEC: The aerodynamic turbine engine code for the analysis of transient and dynamic gas turbine engine system operations, part 1: Model development,'' ASME paper 96-GT-193; ''Part 2: Numerical simulations,'' ASME paper 96-GT-194.
[6] Botros, K. K., Campbell, P. J., and Mah, D. B., 1991, ''Dynamic simulation of compressor station operation including centrifugal compressor and gas turbine,'' J. Eng. Gas Turbines Power, **113**, pp. 300–311.
[7] Botros, K. K., 1994, ''Transient phenomena in compressor stations during surge,'' J. Eng. Gas Turbines Power, **116**, pp. 133–142.
[8] Bemporad, Alberto and Morari, Manfred, 1998, ''Predictive control of constrained hybrid systems,'' International symposium on nonlinear model predictive control: Assessment and future directions, Ascona, Switzerland.
[9] Smeulers, J. P. M, Bouman, W. J., and van Essen, H. A., 1999, ''Model Predictive Control of compressor installations,'' *Proceedings International Conference on Compressors and Their Systems*, London.

Structural Health Monitoring With Piezoelectric Active Sensors

H. A. Winston

F. Sun

B. S. Annigeri

United Technologies Research Center,
411 Silver Lane,
East Hartford, CT 06108

A technology for non-intrusive real-time structural health monitoring using piezoelectric active sensors is presented. The approach is based on monitoring variations of the coupled electromechanical impedance of piezoelectric patches bonded to metallic structures in high-frequency bands. In each of these applications, a single piezoelectric element is used as both an actuator and a sensor. The resulting electromechanical coupling makes the frequency-dependent electric impedance spectrum of the PZT sensor a good mapping of the underlying structure's acoustic signature. Moreover, incipient structural damage can be indicated by deviations of this signature from its original baseline pattern. Unique features of this technology include its high sensitivity to structural damage, non-intrusiveness to the host structure, and low cost of implementation. These features have potential for enabling on-board damage monitoring of critical or inaccessible aerospace structures and components, such as aircraft wing joints, and both internal and external jet engine components. Several exploratory applications will be discussed.
[DOI: 10.1115/1.1365123]

Introduction

Health management of gas turbine engine components requires a set of technologies for detecting and predicting the evolution of damage that results from service loads and operational environments. An important part of this capability is the detection of cracks and other forms of damage in structural components. Traditional non-destructive evaluation (NDE) techniques include eddy current, ultrasonic, fluorescent penetrant, magnetic particle, optical, and radiographic inspection ([1]). These techniques are usually performed off-wing, though some can be applied to accessible on-wing parts. For prognostics and health monitoring of engine components, cracks and other forms of damage must be detected either between periods of engine use or while the engine is operating. Real-time application requires a sensor system that is permanently mounted on or near critical structures such that periodic inspections can be conducted on-line without removing components from the aircraft. The impact and benefit of such on-line health monitoring is that it not only increases flight safety, but also enables converting schedule-based into condition-based maintenance, thus reducing the down time of aircraft and associated maintenance costs.

The piezoelectric active sensor and electric impedance approach has demonstrated unique features that meet the requirements of an on-line health monitoring sensor system. This form of structural health monitoring is based on the use of sensor/actuator, or active sensor elements, to obtain real-time and continuous measurements that reflect the health status of the monitored structure. This paper describes a piezoelectric active sensor-based method that is being developed to detect the presence of cracks and other forms of damage in gas turbine engines and other aerospace structures.

Background and History

A health monitoring technique that measures high-frequency acoustic signatures with piezoelectric active sensors was first reported in ([2]) and later in ([3]). The technique has been tested in a variety of light-weight aerospace structures as well as thick-gauge steel structures, such as a bolted steel joint on a bridge. The technique was found to be very sensitive to incipient structural damage, especially in metallic structures with low damping. Similar work was done by Lichtenwaler et al. [4] in which composite helicopter blades were monitored for delamination by a pair of PZT devices. One device was used as a sensor and the other as an actuator, and the transfer function between them was measured. Structural damage detection by a Wheatstone bridge circuit and a PZT sensor was reported in ([5]). This system eliminated the need for expensive equipment for impedance acquisition. The technology of using coupled impedance measurements for structural health monitoring was summarized by Giurgiutiu and Rogers [6].

This paper reports on recent investigations of the use of high-frequency electromechanical (E/M) impedance spectra for the construction of health monitoring systems designed to detect cracks and other forms of damage in metallic structural components. Piezoelectric lead-zirconate-titanate (PZT) sensors bonded to gas turbine engine components have been used to acquire these spectra. The sensors emit high-frequency ($\sim$100–500 kHz) acoustic excitation whose wavelength is small enough to be sensitive to critical-sized cracks or other small flaws. Early detection of these features can be a key capability of future prognostic health monitoring systems that can support maintenance-on-demand and integrated prognostics and logistics business models.

Principle of Operation

A piezoceramic material, such as PZT, can work as both a sensor and an actuator as a result of the direct and converse piezoelectric effect. When embedded in a structure and driven by an ac voltage, a PZT patch dynamically excites the structure as a result of its field induced strain. The mechanical response of the structure, in turn, modulates the current flowing through the same PZT patch. Suppose a PZT actuator with a mechanical impedance Z_a along the longitudinal direction $\vec{y}$ is clamped at one end, and the other end is connected to a structure with a mechanical impedance Z_s. Using the constitutive equations of PZT material, the coupled electric admittance Y of the PZT between two electrodes in a transverse direction $\vec{z}$ can be described ([7]) by

$$Y = i\omega a\left[\varepsilon(1-i\delta) - \frac{Z_s(\omega)}{Z_s(\omega)+Z_a(\omega)}\frac{(d)^2}{s}\right] \tag{1}$$

Here, δ, ε and d are the dielectric loss tangent, the dielectric constant, and the piezoelectric constant of the PZT material, respectively. s is the mechanical compliance of the PZT, ω is the angular frequency of the driving voltage, and a is a geometric

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Munich, Germany, May 8–11, 2000; Paper 00-GT-051. Manuscript received by IGTI Oct. 1999; final revision received by ASME Headquarters Oct. 2000. Associate Editor: D. Wisler.

constant of the PZT actuator. The first term in Eq. (1) is the capacitive admittance of a free PZT. The second term includes the PZT's own output mechanical impedance as well as the impedance of the structure. For a given PZT sensor integrated into a structure, the coupled electric admittance is then a unique function of the structure's mechanical impedance Z_s and the driving frequency ω. This coupled term appears in an admittance vs. frequency plot as sharp peaks superimposed on a linear capacitive impedance line. Because these peaks correspond to resonances of the structure, they constitute a unique signature of the structure's acoustic behavior. As a result, changes in the impedance signature could indicate structural integrity variations. This coupled electromechanical impedance can be conveniently measured by a commercial electric impedance analyzer.

PZT active sensors and electric impedance measurements for structural health monitoring exhibit high sensitivity to incipient damage, low power consumption, and are relatively non-intrusive to the structure being monitored.

High Sensitivity: The high sensitivity of this technique is the result of two factors. First, the interrogating frequency is ~100–500 kHz. A fundamental requirement for an acoustic signature-based NDE technique is that the wavelength of the acoustic excitation be comparable to the characteristic size of expected forms of structural damage so that such damage can effectively interrupt the transmission and reflection of stress waves. The typical operating frequency of the technique described in this paper is a few hundred kHz. This enables detection of incipient damage on the order of a few millimeters and is possible only when using small low mass transducers.

The second reason for this technique's high-sensitivity is the availability of electric impedance measurement equipment. The electric impedance is much simpler to measure with lower cost and higher accuracy and resolution than its mechanical counterpart. A commercial electric impedance analyzer can have over 100 dB of dynamic range while the mechanical frequency response function obtained with conventional instrumentation, such as accelerometers, force gauges and magnetic shakers, is typically around 60–80 dB. Hence, the mapping of a mechanical characteristic into an electrical analogue greatly facilitates the acquisition of high resolution structural signatures.

Low Power Consumption: The typical drive level for activating PZT sensors is only ~1 V RMS with a maximum current ~10 mA. The corresponding power consumption is in the range of ~10 mW.

Non-Intrusive Sensor: PZT sensor patches are small and have little mass. A typical PZT patch is ~0.75 in.×0.5 in.×0.01 in. and weighs ~0.26 g. As a result, PZT sensors cause little interference with the operation of the structures they are used to monitor.

Several areas of research related to PZT crack detection are reported in the rest of this paper. These include experimental and analytical studies of crack detection by PZT active sensing elements, pattern recognition and temperature correction algorithms, and investigations into the use of these devices on realistic gas turbine engine components.

Experimental Investigation of Sensitivity and Signature-Damage Correlation

To quantify the technique's sensitivity, minimum detectable crack sizes were investigated experimentally. In order to simplify data analyses, steel beams with simple rectangular cross sections were tested under free-free boundary conditions. The beam dimensions were 5.76 in.×1.00 in.×0.125 in. with 0.75 in.×0.5 in.×0.01 in. PZT patches bonded 1 in. from their ends. The beams were hung with thin nylon strands in air, and their impedance spectra from 50 kHz to 350 kHz were measured.

Specimen-to-Specimen Signature Variations. Part damage is identified by variations of impedance signatures from baseline signatures. In order to determine whether generic baseline signatures could be used for different types of components, specimen-to-specimen variations in E/M impedance spectra were studied. Impedance signature differences between beam specimens were found to depend primarily on variations in beam geometry and sensor location.

Beam specimens were made of cold-rolled steel plates with machining tolerances under 0.001 in. PZT patch positions were made as repeatable as possible given that these sensors were manually bonded. Nevertheless, under these conditions, dimensional and sensor location variations were unacceptably large and, hence, damage detection relative to common baseline signatures was not found to be feasible. Figure 1 shows the electric impedance spectra of four specimens having the same nominal size. Beams one and four have a similar signature pattern, as do beams two and three. However, the difference between the two groups is significant and larger than that caused by typical forms of structural damage. Therefore, in this case, damage is only detectable by comparing impedance signatures against the baselines of individual components.

Sensor Re-bonding Repeatability. Investigations into sensor re-bonding variability were conducted. The experiments showed that spectral variations due to re-bonding of the PZT sensor are of the same order of magnitude as those due to moderate amounts of damage. This result implies that PZT sensors must be permanently embedded in or bonded to the structures they are used to monitor. In particular, to ensure repeatable signature readings, they cannot be temporarily attached to a structure whenever measurements are needed.

Correlation of Signature With Damage Location. The possibility of determining crack locations from impedance signatures was supported by experiments that measured how the location of small drill indents affected high-frequency resonant spectra. Indents were made in a free-free supported beam at increasing distances from a PZT sensor, and each indent location was shown to correlate with a negative frequency shift of a different resonance peak. Figure 2 shows the results of one such experiment in which two indents were drilled in a specimen and correlated with significant shifts in the far right peak for indent 1 and significant shifts in the far left peak for indent 3.

Effect of Sensor Material on Signature. Different sensor materials were tested to study their impact on damage sensitivity and temperature stability. The effect of piezoelectric material on the impedance signature was also investigated, and no noticeable variation was observed between hard piezoelectric (PZT 4) and

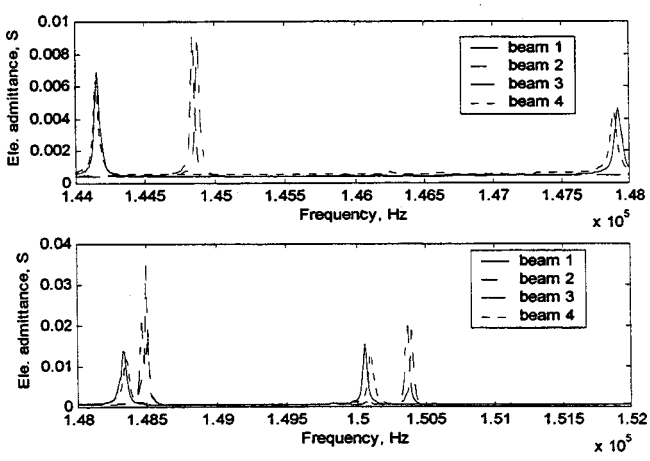


Fig. 1 Impedance variations between identical beam specimens

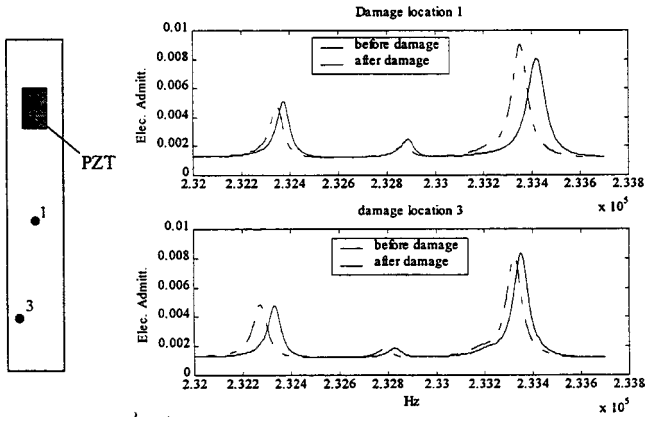


Fig. 2 Impedance signature variation with damage location (solid line: before damage; dashed line: after damage)

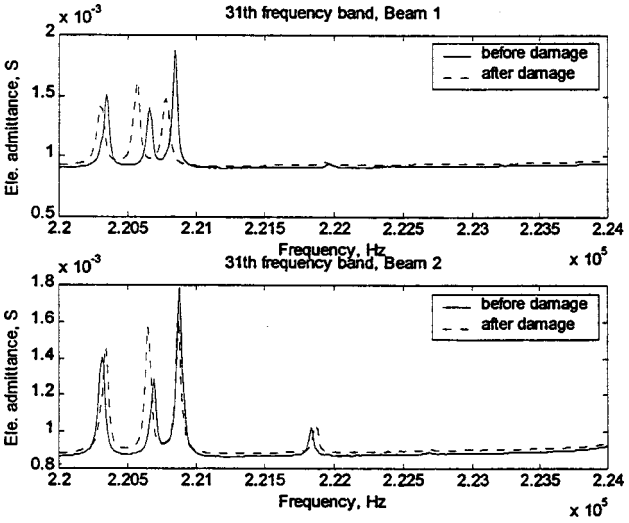


Fig. 3 Signature changes caused by 0.060 in. EDM cracks

soft piezoelectric (PZT 5A) sensors. However, hard PZT sensors demonstrated less temperature dependence than soft PZT sensors on the same beam specimens.

Effect of Temperature on Signature. It is well established that environmental temperature has a significant effect on electromechanical impedance signatures. Thus, an appropriate temperature correction must be employed before structural damage can be identified by an impedance signature change. A cross correlation-based algorithm was implemented to address this issue, and the details will be discussed in a later section of this paper. In addition, a dual PZT patch configuration was studied as a hardware alternative for temperature correction. This technique uses one patch as a sensor and another as an actuator and measures the transfer function between them, instead of observing the point transfer function at one location with a single sensor. The dual PZT patch configuration demonstrated less spectrum variation with sensor temperature change than single PZT patches, without compromising detection sensitivity. However, this method is harder to apply because it requires an additional sensor and associated instrumentation.

Signature Sensitivity to Damage. Cracks in metallic structures are not easy to generate under normal laboratory conditions within short periods of time. Moreover, additional structural perturbations could occur during crack initiation and propagation that could contaminate controlled tests. To accelerate testing of PZT crack sensitivity, while strictly controlling test conditions, beam specimen cracks were simulated by electric discharge machine (EDM) cutting. A small 0.009 in. wire gauge was used, and the width of the resulting simulated cracks were about 0.012 in. This technique was used to study the effect on impedance signatures of cracks at different locations, orientations and sizes.

Figure 3 illustrates signature pattern changes before and after 0.060 in. long EDM cuts were made on the beam specimens. Although identical EDM notches were applied to two of these specimens, the impact of these notches on each specimen's signature was different. This is believed to be due to specimen variability—the baseline signatures of beams can differ significantly within a given frequency band. Therefore, over this interval of frequencies, one specimen might be more sensitive to damage than another. It was also observed that the smallest crack that can be unambiguously detected by an impedance signature change is approximately 0.060 in. in length, when the crack orientation is perpendicular to the beam's longitudinal axis and located 2 inches from the sensor.

Table 1 quantifies these results with the damage identification algorithm described below in Eq. (2). The relative damage index was calculated over a frequency band from 200 kHz to 255 kHz, after temperature effects were corrected. The index for a crack

Table 1 Relative damage indices caused by 0.06 in. and 0.12 in. cracks

Damage Index Caused by 0.06" Crack			Damage Index Caused by 0.12" Crack		
Distance from Sensor	Perpen-dicular	Parallel	Distance from Sensor	Perpen-dicular	Parallel
1"	1	0.35	1"	1.72	0.52
2"	0.66	0.20	2"	1.05	0.32

0.060 in. long, perpendicular to the beam's longitudinal dimension, and 1 inch from the sensor is defined to be unity. Although there is no quantitative relationship between values of this scalar damage index and crack location or orientation, a positive correlation nevertheless exists between index values and the extent of structural damage. A crack parallel to the sensor's longitudinal dimension produces a smaller signature variation, as does a crack 2 inches from the sensor. A crack 0.12 inches long causes a larger signature variation than a 0.060 inch crack under the same conditions. Therefore, this simple statistical approach can be used for approximate estimations of damage levels.

Finite Element Models

In order to identify specific forms of damage by impedance signature changes, it is necessary to establish improved correlation between acoustic signature changes and changes in a component's structural integrity. In addition, to develop robust pattern recognition algorithms that can classify situations other than those seen in the laboratory, a high-frequency structural model is needed to study the signatures associated with additional types of damage. For this purpose, a finite element model (FEM) was developed of the same free-free steel beam and PZT sensor configuration used in the above experimental investigations.

An initial step toward analytical crack prediction was made by verifying finite element model predictions at frequencies approaching those typically used in the impedance spectrum approach, but much higher than those normally studied in structural dynamic analyses. The objective was to validate this model by comparing analytical mode shapes and frequency responses with experimental ones. Predictions for natural frequencies of the beam up to 100 kHz were obtained. The effect of a PZT sensor on the predicted modal response of a beam was also studied. Good agreement was also found for these modes with the laser Vibrometer technique described below.

Effect of PZT Sensor on the Modal Response. The effect of incorporating a PZT sensor model on the modal shapes and frequencies was investigated. The PZT was modeled with 64 brick elements that had a total thickness of 0.01 in. This analysis was performed in the frequency range from 0 Hz to 60 kHz and included a total of 45 modes. The results for modes 20 to 25 are listed in Table 2. The percent difference errors for the other 39 modes were similar to the values of the six reported modes. As shown in the table, adding the PZT sensor to the beam model does not significantly affect the frequencies of high-order modes. Therefore, it was possible to simplify the finite element model by omitting the PZT sensor.

Experimental Validation of Mode Shapes and Frequencies. The mode shapes from the finite element model were validated by comparing them with experimental data. Collaborators at North Carolina A&T University conducted a series of measurements on the modal shapes of our canonical steel beam with a Scanning Doppler Laser Vibrometer (SDLV) system. This system is able to analyze structural vibrations up to 250 kHz with a sensitivity of 0.5 μm/s. Complete laser vibrometer studies of cantilever beam mode shapes and associated resonant frequencies were obtained for frequencies up to 200 kHz. Good agreement was obtained between the SDLV data and the above finite element results. Figure 4 illustrates corresponding beam mode shapes at 45,773 Hz, as derived from the finite element model (top), and at 45,732 Hz, as derived from the SDLV (bottom).

Table 2 Effect of PZT model on modal frequencies

Mode No.	Mesh w/o PZT (Hz.)	Mesh with PZT (Hz.)	Percent Difference
Mode 20	25879	25926	-0.18
Mode 21	26547	26611	-0.24
Mode 22	27693	27693	0.00
Mode 23	27979	27979	0.00
Mode 24	30144	30160	-0.05
Mode 25	30590	30605	-0.05

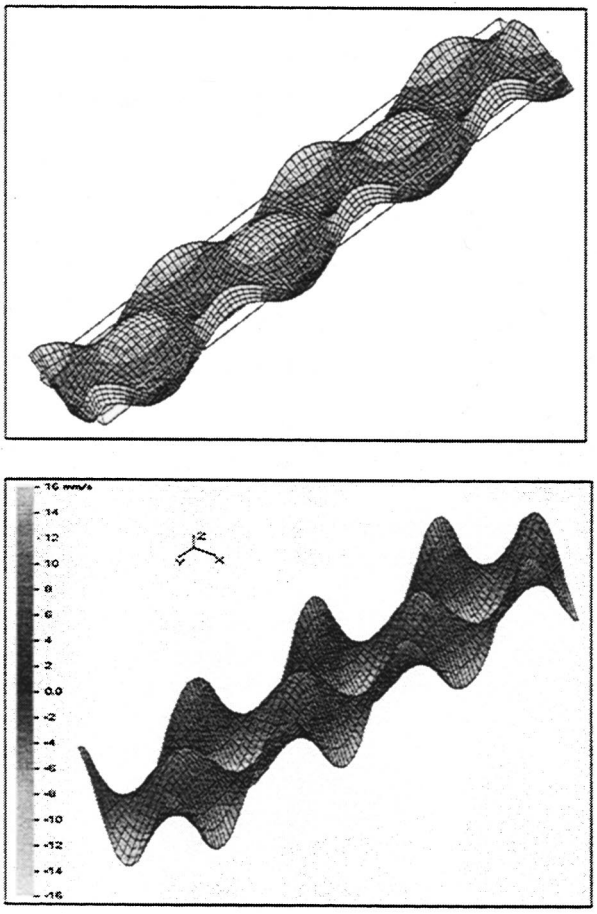


Fig. 4 Modal shape of the beam, top by FEM (45,773 Hz), bottom by SDLV (45,732 Hz)

Data Processing and Damage Identification

Large amounts of high-frequency impedance spectrum data were obtained from each sample/condition pair. As a result, algorithms for expediting data exploration, analysis, and reduction were developed. These codes can extract resonance peak features, correlate these features between sets of similarly prepared specimens, and use them to detect features for distinguishing the effects of cracks and other defects. Impedance spectra consist of resonance peaks superimposed on a slowly varying background. Structural integrity variations can result in resonance peak shifts, splits (e.g., from the splitting of degenerate vibration modes in symmetric components), merges or the appearance of new peaks. Two types of algorithms were developed to automate the detection and measurement of these variations.

Simple free-free beam specimens exhibited clearly separated resonance peaks. As a result, local features of these signatures could be extracted and used to compare multiple spectra. For this purpose, a three-stage pattern recognition algorithm was developed. First, a peak detection algorithm locates resonance peaks and catalogues their frequencies, amplitudes, and widths. Second, a list of corresponding peaks in two impedance spectra (e.g., from nominal and damaged specimens) is generated. This procedure uses a metric that measures the alignment (in frequency domain) of peaks from two spectra that have similar amplitudes and widths. Finally, a routine was developed that accepted a list of corresponding peaks from two or more spectra and returned mean, standard deviation, and coefficient of variation statistics about corresponding peak locations, amplitudes, and widths.

Engine component specimens exhibited complex impedance spectra without clear indications of individual resonance peaks. As a result, only global features of these signatures could be measured and used to compare multiple spectra. For this purpose, a statistical delta (SD) approach that measures the dissimilarity between two impedance signatures was used. For signatures consisting of n discrete impedance measurements Y^i, taken over a frequency range of interest, the SD metric is defined as:

$$\mathrm{SD} \equiv \sum_{i=1}^{n} \frac{\sqrt{(Y^i_{\mathrm{damage}} - Y^i_{\mathrm{baseline}})^2}}{\Delta Y^i_{\mathrm{baseline}}} \tag{2}$$

Here, $\Delta Y^i_{\mathrm{baseline}}$ is the maximum baseline signature change due to sources of noise other than structural variations.

Temperature Correction by Cross-Correlation Analysis. Electro-mechanical impedance signatures are significantly affected by environmental temperature variations that alter the material properties of structures and the dielectric properties of sensors. This hampers the utility of these signatures for structural health monitoring. Experiments demonstrate that this temperature variation may result in impedance spectrum changes (e.g., resonance peak shifts) that are comparable in magnitude to those that result from moderate structural damage. Hence, temperature-induced changes might result in unacceptably high false alarm error rates. Methods of physical compensation by secondary temperature sensors and associated electrical circuits would, themselves, be vulnerable to temperature effects and are, thus, not a completely effective means for temperature compensation.

A software-based method was developed to correct impedance spectrum variations caused by thermal effects. The technique is based on differences between spectral changes due to structural

damage and changes due to temperature variations. Within certain frequency bands, temperature-induced changes result in nearly uniform vertical and horizontal translations of impedance spectra, while impedance variations due to structural changes produce relatively localized effects on E/M impedance spectra (e.g., resonant peak splitting and peak-dependent resonant frequency shifts). This distinction enables the removal of the thermal effect by the simple cross-correlation algorithm described in ([7]).

The approach measures the extent to which an impedance spectrum curve is horizontally and vertically translated due to temperature effects. Temperature effects that result in vertical displacements of impedance spectra are compensated by computing and removing the difference between the mean values of the baseline spectrum *X* and the current spectrum *Y*. On the other hand, the cross-correlation between these signals is used to determine how much horizontal translation of spectrum *Y* with respect to baseline spectrum *X* can be attributed to temperature effects. This approach is effective when component damage results in minor and localized perturbations in impedance signatures.

Figure 5 illustrates the cross-correlation method of temperature compensation. The top, or first, plot shows a baseline impedance signature before and after a thermal shift. The second shows the result of applying a thermal correction. The third shows the thermally corrected signature of a damaged specimen versus the baseline. The bottom plot shows the cross-correlation coefficient corresponding to the horizontal thermal shift of the spectrum and how the correction is determined. The circle on the curve indicates the peak location (780 data points) of the cross-correlation coefficient, if there is no thermal shift. Due to the thermal shift, the peak is actually moved to the asterisk. Moving the new impedance signature by the distance between these two locations, with respect to the baseline, yields thermally corrected results that retain the damage variations, as shown in the third plot.

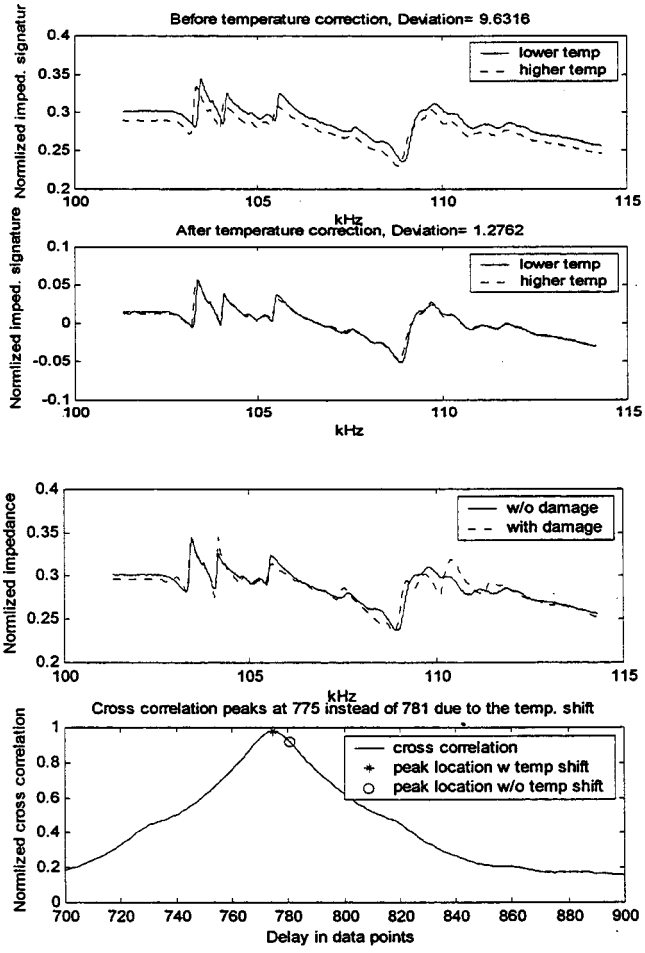


Fig. 5 Temperature compensation by cross correlation analysis

Engine Component Studies

In addition to the simple cantilever beam experiments reported above, we have studied the detection of cracks with high-frequency impedance spectra in several jet engine components. These components were selected in order to investigate the feasibility of using E/M impedance signatures for structural health monitoring of realistic components with complex geometrical shapes. The specific parts used in these studies were selected on the basis of sample availability and part accessibility.

Isolated Component. Cracks measuring 5 mm×0.5 mm×0.5 mm inscribed on engine airfoils by a Dremel tool were found to be consistently detectable with PZT sensors. However, significant specimen-to-specimen variability implies that damage is best assessed by tracking changes in the spectra of individual components.

In Situ Component. In order to mature this technology and better gauge downstream commercialization risks, PZT prognostic sensors were included in a seeded-fault engine test (SFET). The goals of this test were to (1) evaluate the PZT sensor's survivability in a harsh engine environment (i.e., determine the robustness of the sensor and adhesive to high temperatures and extreme vibration levels), (2) determine the repeatability of the high-frequency impedance measurements during and after engine cycling, and (3) evaluate the sensitivity of the sensor to seeded engine faults.

Five PZT patches were mounted on the external surface of the Nozzle Static Structure (NSS) of a retired F100 engine. This location was selected because of its accessibility, susceptibility to significant noise and vibration, and exposure to moderately high temperatures. (PZT sensors lose their piezoelectric properties above a characteristic Curie temperature.) Broadband impedance signatures were acquired from these sensors when the engine was shut down and while it was operating in idle, intermediate, and augmentation modes. The measurements demonstrated good repeatability over a 100 kHz frequency window, despite the presence of strong grounding noise and electromagnetic interference from other instrumentation cables. As expected, the signature patterns shifted as the temperature of the NSS increased. The patterns subsequently returned to their baseline values when the engine cooled to ambient temperatures.

All the sensors survived a two week engine test during which about twenty 90 minute engine cycles (idle, intermediate, augmentation, and stop) were executed. No cracks in or performance degradation of the sensors were observed. [During the second week of the SFET, one sensor was partially debonded from the NSS. This was subsequently found to have resulted from defective bonding (i.e., an air bubble was trapped in the adhesive) and water intrusion into the bonding layer. The sensor was re-bonded, and good signatures were obtained for the remainder of the SFET.]

After verifying sensor survivability and data repeatability, five artificial cracks were made in the externally accessible NSS to simulate structural damage on an engine component. Noticeable signature changes occurred when a 3/8 inch crack was made in the NSS 2.5 in. from one of the sensors. Only a limited opportunity for experimentation existed during the SFET. As a result, this defect size was selected to be small enough to be of interest to applications engineers, but large enough to assure initial detectability during the SFET. It is not known whether smaller cracks could have been detected and, if not, whether 3/8 inch cracks are at the theoretical lower limit of detectability. The variation from baseline increased significantly when the crack was subsequently expanded to 1 inch.

In summary, three conclusions can be drawn from the results of the SFET. The PZT sensors can survive an engine environment without failure or noticeable degradation in performance. The E/M impedance signatures demonstrated good repeatability, even in the presence of strong electromagnetic interference and grounding noise. Induced cracks ranging in size from 3/8 inch to 1 inch were reliably detected on nozzle static structures. Hence, sensor survivability and measurement repeatability risks were significantly addressed by the SFET. Although damage sensitivity was demonstrated, work remains to be done on discriminating damage effects from other potential sources of spectrum variations.

Summary

This paper has reviewed investigations into the use of PZT active sensors for gas turbine engine component damage detection. A technique for automated damage detection with PZT patches has been demonstrated based on global spectrum changes measured by a sum of squared impedance differences over a range of discrete frequencies. Correlation of impedance spectrum responses with the location and size of various forms of damage was also experimentally investigated, and a novel approach for temperature correction of impedance spectra was introduced. Uses of this technology with representative gas turbine engine components were presented. A finite element model at high frequencies was also developed which shows that the dynamics of a cantilever beam can be accurately predicted at frequencies typically used in the impedance approach. This has paved the way for analytically predicting impedance spectrum changes caused by structural damage.

Acknowledgments

The authors would like to express their thanks to Mr. Doug Gagne for conducting beam finite element modeling, Dr. Mark Schulz (NCA&T University) for performing the experimental SDLV investigations, and Steve Budrow and Wayne Layman (Pratt and Whitney) for providing logistical support during the SFET tests. The authors thank Dr. Alan Eckbreth, Dr. Ed Rogan, Mr. Frank D. Gass (Pratt and Whitney), and Dr. Sharayu Tulpule for supporting this research effort. The authors would also like to express special thanks to Dr. Zaffir Chaudhry (Memry Co.) for helping to initiate and contributing to this research.

References

[1] Bray, D. E., and Roderick, S. K., 1989, *Nondestructive Evaluation*, McGraw-Hill, New York.
[2] Sun, F., Chaudhry, Z., Liang, C., and Rogers, C. A., 1994, ''Truss Structure Integrity Identification Using PZT Sensor-Actuator,'' *Proceedings of the Second International Conference on Intelligent Materials*,'' June 5–8, Colonial Williamsburg, VA, pp. 1210–1222.
[3] Chaudhry, Z., Sun, F., and Rogers, C. A., 1995, ''Localized Health Monitoring of Aircraft via Piezoelectric Actuator/Sensor Patches,'' *Proceedings, SPIE North American Conference on Smart Structures and Materials*, San Diego, CA.
[4] Lichtenwaler, P. F., Dunne, J. P., Becker, R. S., and Baumann, E. W., 1997, ''Active Damage Interrogation System for Structural Health Monitoring,'' Proc. SPIE, **3044**, pp. 186–194.
[5] de Vera Pardo, C., and Guemes, J. A., 1997, ''Embedded Self-Sensing Piezoelectric for Damage Detection,'' *Structural Health Monitoring—Current Status and Perspective*, Proceedings of the International Workshop on Structural Health Monitoring, Stanford University, Stanford, CA, Sept. 18–20.
[6] Giurgiutiu, V., and Rogers, C. A., 1997, ''Electro-Mechanical (E/M) Impedance Method for Structural Health Monitoring and Non-Destructive Evaluation,'' *Structural Health Monitoring—Current Status and Perspective*, Proceedings of the International Workshop on Structural Health Monitoring, Stanford University, Stanford, CA, Sept. 18–20.
[7] Sun, F., Chaudhry, Z., and Rogers, C. A., 1995, ''Automated Real-Time Structural Health Monitoring via Signature Pattern Recognition,'' SPIE North American Conference on Smart Structures and Materials, San Diego, CA.

A Fiber-Optic Tip-Shroud Deflection Measurement System

R. A. Rooth

W. Hiemstra

KEMA Nederland B.V.,
P.O. Box 9035,
6800 ET Arnhem, The Netherlands

A Dutch utility faced the fact that a second stage set of turbine blades of a gas turbine had to be replaced long before the estimated lifetime as a result of tip shroud deflection. This deflection caused the risk of loss of support between the individual tip shroud segments. The goal of this paper is to find the cause of the problem, to see how it increases with time, and to take appropriate action to prevent the problem from occurring again. A fiber-optic tip-shroud deflection monitor has been developed and tested on a gas turbine to study this phenomenon in real time. The optical system is based on astigmatism to derive distance information from the measured optical signals. Characteristic features from the system are the good spatial resolution of about 1 mm, the distance resolution of about 0.1 mm and the distance from the probe tip to the target of about 20 mm. These specifications are difficult to achieve with, for example, capacitive sensors. The probe tip can with stand temperatures of about 500°C. The system can be calibrated in situ, given a stable operation of the gas turbine. This is accomplished by stepping the probe tip over some distance and recording the signals corresponding to a certain point on the tip shroud. The instrument has been used to monitor the tip-shroud deflection in a gas turbine at various loads and over a time span of several months. The results indicate that the deflection can be divided in a part depending on the load and a part that is a permanent deflection. Based on the results, it can be judged whether blades need to be rejected because of a too large deflection. [DOI: 10.1115/1.1364498]

Introduction

One of the Dutch utilities faced the fact that a second stage set of turbine blades had to be replaced long before the estimated lifetime as a result of tip shroud deflection. This deflection caused the risk of loss of support between the individual tip shroud segments. There was a need to understand the nature of the problem. If a correlation could be developed between operation of the gas turbine and the development of the deflection then there would be an outlook into the possibility of predicting the deflection which could lead to cost savings related to blade replacement.

Different systems have been described in the literature, optical triangulation (Barranger [1]) or astigmatism (Davinson [2,3]), capacitive (e.g., Chivers [4], Müller [5]), microwave (Grzybowski [6]). All have their advantages and drawbacks. The optical triangulation system is insensitive to all kind of environmental influences, but bulky and expensive. The capacitive sensors find wide use, but have poor spatial resolution or low sensitivity at large measurement distance. They rely on a frequency shift of an oscillating signal in a situation when a blade passes as compared to the situation when no blade tip is present. This makes them unsuited for shrouded blade rows. The microwave technology is fairly new. The astigmatic sensor head can be made compact, has good sensitivity at large measurement distances and an extremely good spatial resolution. Diode laser technology makes the ground station compact as well. Because of these properties the astigmatic sensor was chosen for the measurement problem at hand.

Measurement Principle

In order to understand the starting points for the development of a sensor to measure the deflection in real time it is necessary to understand the deflection problem. This problem is shown in Fig. 1, which is taken from a technical information letter (Ref. [7]). It can be seen that the second stage blades support each other through the contact between the individual tip shroud segments. If the blades start to deflect or ''bend'' then the contact surface reduces which eventually leads to rejection of the blades because of the risk of failure of the row. The measurement problem focuses on the measurement of the step size at the contact of the two segments. This step could be measured using conventional capacitance type sensors but there are two problems. First, the tip shrouds feature labyrinth sealing, which is accomplished by two strips on top of the shroud. This prevents the sensor of being placed close to the surface of the shroud, this would result in poor sensitivity of the sensor unless it would have a large diameter. Second, to have a good measurement of the step size, an excellent spatial resolution is required. To overcome these, an optical approach has been chosen which is made possible by advances in fiber-optic and laser technology. The probe of the measurement system is based on optical astigmatism, see Fig. 2. This principle has been published by (Davinson [2]) but to our knowledge there is no published use of this technique on a gas turbine. Using this

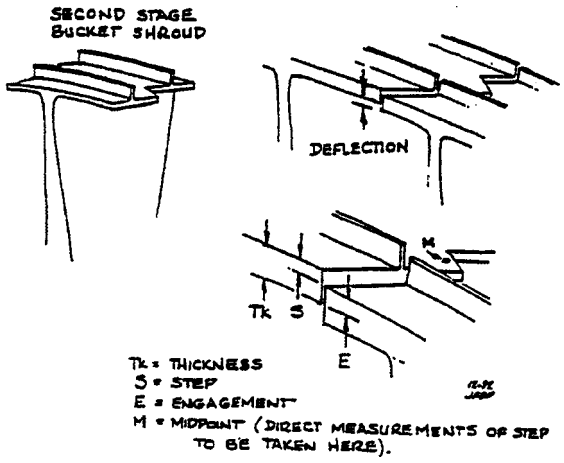


Fig. 1 Tip shroud deflection

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Munich, Germany, May 8–11, 2000; Paper 00-GT-053. Manuscript received by IGTI Oct. 1999; final revision received by ASME Headquarters Oct. 2000. Associate Editor: D. Wisler.

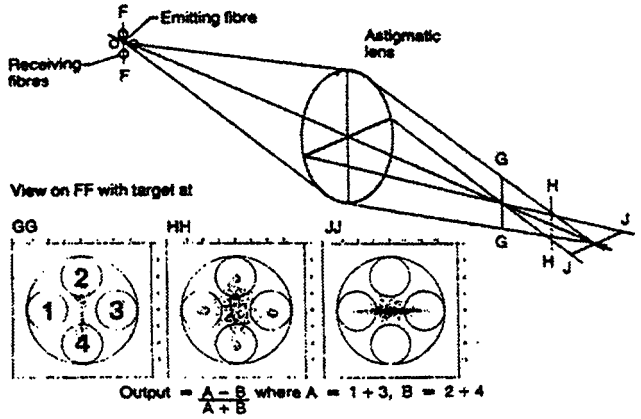


Fig. 2 Principle of the astigmatic optic deflection probe

principle the system can measure distances at about 20 mm from the probe tip with a spatial resolution of about 1 mm.

System Configuration

Probe Optics. Laboratory experiments have shown that the best configuration for the optical system is an astigmatic objective consisting of a planoconvex spherical and a planoconvex cylindrical lens. Through the center of the lenses a small hole has been drilled through which a laser beam is focused onto the target. This configuration is shown in Fig. 3.

The average image distance of the astigmatic optic is about 100 mm when the object distance is 22 mm. This results in a big enough image spot at the four fibers that create a four quadrant detector, see Fig. 2. The fibers have a 400 μm diameter. A varying distance of the target around the 22 mm distance from the probe tip results in an image shape that varies from a horizontal ellipse, through a circle to a vertical ellipse.

Mechanical construction. The probe, mounted on the gas turbine, is shown in Fig. 4. All components in the probe are capable of handling a temperature of 600°C. There is one exception, which is the weld of the gold-coated fiber within the pressure feedthrough. This feedthrough can withstand a pressure of 100 bar and a temperature of 200°C. This feedthrough is located 15 cm outside the gas turbine. The local temperature there is about 100°C. Using a dummy probe, equipped with thermocouples, the temperatures at different points in the probe were determined. It was determined that there were no points where the temperature exceeded the acceptable values. The gold-coated fibers in the hot end of the probe are protected by a 0.7 mm diameter stainless steel capillary. Their outer end is mounted in a 2 mm diameter ferrule that slips into a mounting plate. The entire probe is mounted on a flange that used NPT thread to fix it to the turbine. Using a spring, the probe is pressed against a reference surface. Cooling air from the compressor is used to keep the probe at an acceptable temperature. The construction is mainly from grade 310 stainless steel.

Signal Processing. Figure 5 shows the layout of the cabinet

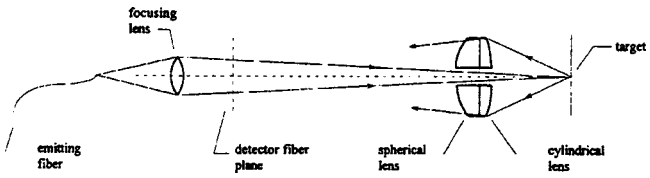


Fig. 3 Optical configuration of the astigmatic probe. Spherical lens $f=25$ mm. Cylindrical lens $f=83$ mm.

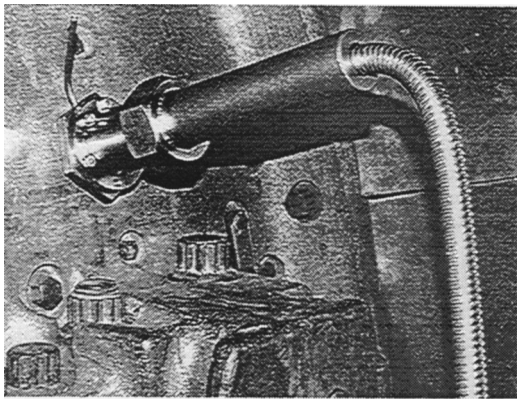


Fig. 4 Tip shroud deflection probe, mounted on the turbine

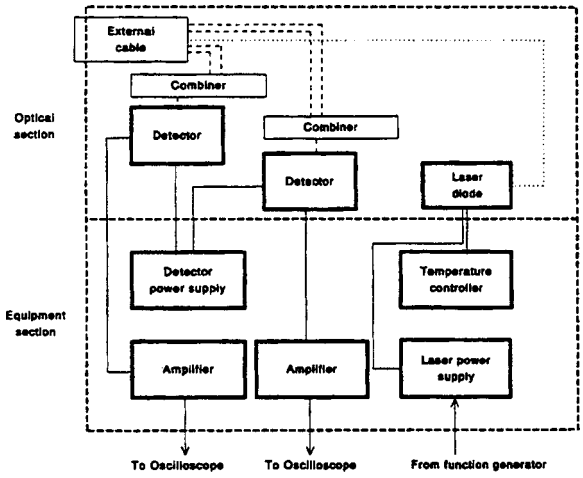


Fig. 5 Layout of the detection cabinet

that houses the fiber optics, detectors, and laser system. Signals from the detector fibers are first guided to the detection cabinet outside the gas turbine package. This is done by tapered fibers, reducing the fiber diameter from 400 to 100 μm. Then they are combined using bifurcated fiber combiners and fed into the two silicon detectors. The signals are amplified and then directed to a digital, averaging, storage oscilloscope. Triggering of the system is accomplished by use of the key-phasor signal. Illumination of the target is accomplished by a near infrared diode laser having a maximum power of 100 mW. This laser is sine modulated to avoid problems with drift of the detectors and possible stray light. The modulation frequency is 333 kHz. When the signals are acquired by the oscilloscope, they are processed by the controlling PC. Every period of the modulation is reduced to a single ''distance'' signal point by determination of the amplitude of the signal in that period. As a result we have two arrays of signals (A and B). Each successive data point corresponds to a measurement at a distance of 1.2 mm from the previous point (tangential distance).

Calibration

As is shown in Fig. 2, the output of the sensor can be described in terms of $(A-B)/(A+B)$, where $A=1+3$ and $B=2+4$. The numbers here refer to the optical radiation signal collected by a fiber with that number. Typical calibration curves are shown in Fig. 6.

The calibration curves form the relation between the measured signals and the actual distance from the target to the probe tip. These curves are stored in the computer and used in the distance

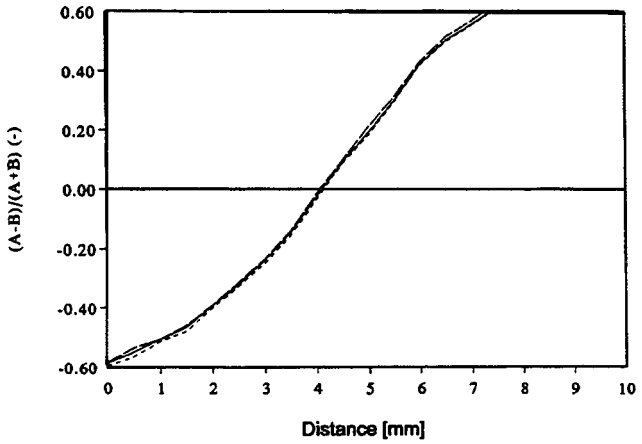


Fig. 6 A series of three successive calibration runs, showing the good reproducibility

calculations. The normalization of the signal (by the division by $A+B$) results in a calibration that is hardly depending on the reflection of the target or the laser power.

Fiber-optic Deflection Monitor Specification

The resulting deflection sensor has the specifications listed in Table 1.

Result

Measurements. The system has been used to monitor the tip shroud deflection of the second stage of a gas turbine at different loads of the engine in order to see whether a correlation between the deflection and turbine power exists. Figure 7 shows an example of the recorded distances with the system. One quarter of a revolution is shown. The distance scale is in relative units as the interest is in the distance difference from one shroud segment to another. As can be seen, the distance rises sharply and then decreases gradually. This means that the tip shrouds, and possibly also the blades, are leaning ''forward'' which can be expected with a turbine under load. An interesting feature can be seen at about one third of the trace corresponding to a single shroud segment. There is a disturbance in the distance signal. This is caused by the passage of a cooling hole underneath the measurement spot. At blade number 15 in the graph this disturbance is very large. This is caused by an enlargement of the hole for identification purposes. This can be seen in Fig. 8.

Table 1 Tip shroud deflection monitor specifications

Spatial resolution	1 mm
Distance resolution (after several minutes of measurement time)	0.1 mm
Free distance from probe tip	20 mm
Maximum probe temperature	500°C
Maximum pressure	8 bar

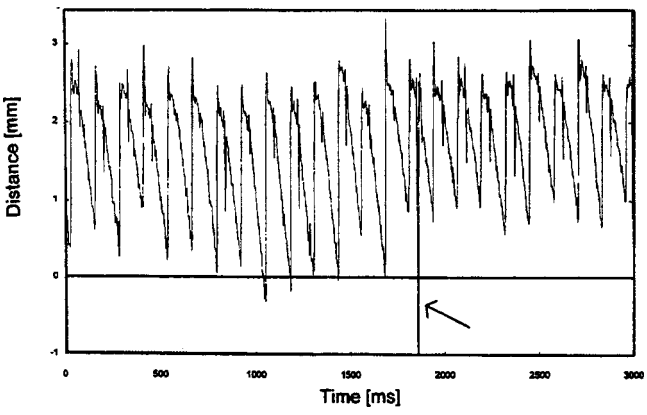


Fig. 7 Example of the distance signal of the tip shroud monitor, the arrow shows the measurement at the enlarged cooling hole position

Fig. 8 Picture of the tip shroud segments. The picture shows the cooling holes and the enlarged hole for identification of the blades.

As we are mainly interested in the ''steps'' occurring at the transition from one shroud segment to the other, it is possible to reduce the distance measurement set to 92 steps. Such sets are shown in Figs. 9–11.

The mechanical measurements are correlated to the optical measurements by means of the tip shroud with the enlarged cooling channel outlet, visible in Fig. 8.

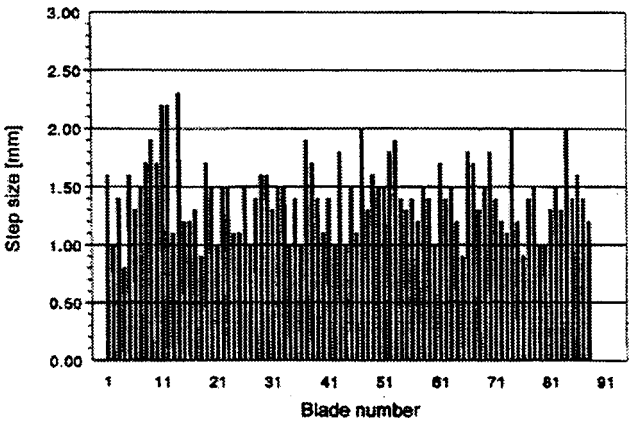


Fig. 9 Step sizes with turbine on turning gear, 1200 rev/min

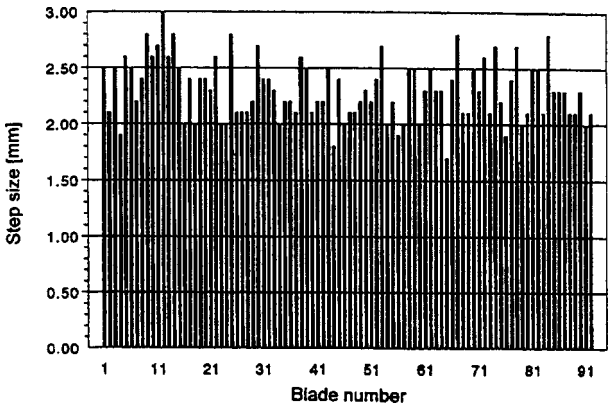


Fig. 10 Step sizes with turbine at full load, 36 MW. Turbine frequency 85 Hz.

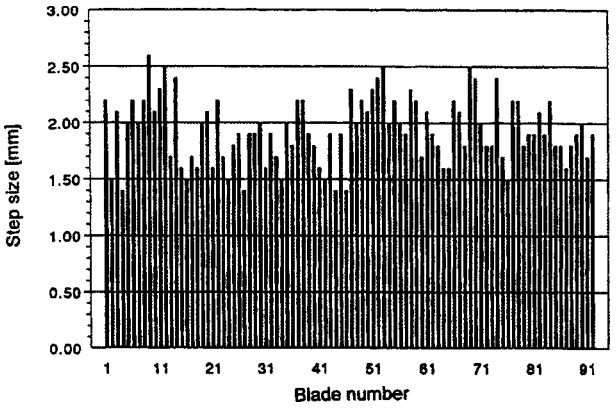


Fig. 11 Step sizes with turbine at rest, mechanical measurement during an overhaul

Interpretation of Findings. The results show the following.

1 The distance step from a certain tip shroud segment to the next is not the same from one start to another.
2 The average step size is larger at larger loads (2.3 mm at 36 MW versus 2.0 mm at 14 MW)
3 On turning gear the average step size is smaller than at rest (1.4 mm versus 1.9 mm).

Ad 1. This is probably caused by the fact that the blades are loosely mounted in the disk. This means that at a start all blades more or less ''tumble'' over when the engine switches from turning gear to load. This will not occur at exactly the same time for all the blades.

Ad 2. The increased average step size at larger loads suggests that there is an elastic contribution to the deflection.

Ad 3. As a result of the loose mounting, on turning gear the blades are inclined backwards. The steps with the engine at rest are all measured after pressing the blades in the direction of rotation.

These results show the importance of the on-line measurements. The tip shroud deflection under full load conditions is more severe than with the turbine at rest. It is the deflection under load that is the cause of possible problems.

On-line Calibration of the System

The measurement campaigns on the gas turbine have shown that it is desirable to increase the stability of the calibration as the high temperature of the probe gives rise to small distortions. It would be nice if the system could be calibrated in situ. This has been realized by mounting the probe on a translation mechanism. Under stable operating conditions, the probe can be traversed in such a way that the distance to the target changes in a controllable way. A certain point on one of the tip shroud segments is taken as a reference point. This point is used for obtaining the distance signal related to a certain setting of the translation mechanism and hence of the probe tip. A calibration curve can then be constructed by moving the probe, e.g., 0.5 mm stepwise and recording the distance signal.

Conclusions. A fiber-optic tip-shroud deflection monitor has been developed and tested on a gas turbine. The optical system is based on astigmatism to derive distance information from the measured optical signals. Characteristic features from the system are the good spatial resolution of about 1 mm, the distance resolution of about 0.1 mm and the distance from the probe tip to the target of about 20 mm. The probe tip can withstand temperatures of about 500°C. The system can be calibrated *in situ*, given a stable operation of the gas turbine. The instrument has been used to monitor the tip-shroud deflection in a gas turbine at various loads and over a time span of several months. The results indicate that the deflection can be divided in a part depending on the load and a part that is a permanent deflection. Based on the results, it can be judged whether blades need to be rejected because of excessive deflection.

Acknowledgment

The author expresses his gratitude to the Collective Order Production of the Dutch utilities for funding this research as well as EPON and Thomassen Int. for providing the possibility to test the system on one of their gas turbines.

References

[1] Barranger, J. P., and Ford, M. J., 1981, ''Laser Optical Blade Tip Clearance Measurement System,'' J. Eng. Power, **103**, No. 2, p. 58.
[2] Davinson, I., 1987, ''Astigmatic Turbine Blade Tip Clearance Sensor for Gas Turbine Aero Engines,'' Proc. SPIE, **863**, pp. 60–63.
[3] Davinson, I., 1990, ''The Use of Optical Sensors and Signal Processing in Gas Turbine Engines,'' Proc. SPIE, **1374**, pp. 251–265.
[4] Chivers, J. W. H., 1989, ''A Technique for the Measurement of Blade Tip Clearance in a Gas Turbine,'' 25th Joint Propulsion Conference, AIAA Paper No. 89-2916.
[5] Müller, D., et al., 1996, ''Capacitive Measurement of Compressor and Turbine Blade Tip to Casing Running Clearance,'' ASME Paper No. 96-GT-349.
[6] Grzybowski et al., 1996, ''Microwave Blade Tip Clearance Measurement System,'' ASME Paper No. 96-GT-002.
[7] Exxon, Peterson, J. F. D., 1993, notes of the 1992 Frame 6 Gas Turbine User's Roundtable.

Gas Turbine Compressor Washing State of the Art: Field Experiences[1]

J.-P. Stalder
Turbotect Ltd., P. O. Box 1411
Ch-540 Baden, Switzerland

Technology development in gas turbine compressor washing over the last 10 years and today's state of the art technology is presented in this paper. Based on various long term field tests and observations, correlation between rate of power degradation and atmospheric conditions can be established. Questions about compressor on line washing with water alone against the use of detergents, as well as washing frequencies are also addressed in this paper. Performance degradation behavior between gas turbines of different sizes and models can be explained with an index of sensitivity to fouling. The implementation of an optimized regime of on line and off line washing in the preventive turbine maintenance program is important. It will improve the plant profitability by reducing the costs of energy production and contribute to a cleaner environment.
[DOI: 10.1115/1.1361108]

Introduction

Gas turbine (GT) cleaning was made in the early days by crank soak washing and/or by injecting solid compounds such as nutshells or rice husks at full speed with the unit on line. This method of on line cleaning by soft erosion has mainly been replaced by wet cleaning since the introduction of coated compressor blades for pitting corrosion protection. Further, unburnt solid cleaning compounds and ashes may also cause blockage of sophisticated turbine blade cooling systems if ingressed into the GT air cooling stream. At the beginning of the introduction of compressor wet cleaning in the 1980's, time intervals between on line washing and the combination with off line washing had to be established. Further, there was also a belief among many users that on line washing could replace off line washing. Hoeft [1] quoted that an airflow reduced by 5 percent due to fouled compressor blades will reduce output by 13 percent and increase heat rate by 5.5 percent. With today's large scale use of gas turbines in combined cycle base load application as well as with their increase in nominal output, gas turbine compressor washing has gained more and more attention by their owners.

Combination and Frequency of On Line and Off Line Washing

A first long term test on the combined effect of on line and off line washing at various washing intervals was performed over 4000 operating hours in the mid 1980's at the Energieproduktiebedrijf Utrecht (UNA) PEGUS 100 MW combined cycle plant in The Netherlands (two gas turbines of 30.7 MW site output and one 38.6 MW steam turbine.) The plant is situated beside the Merwedekanaal on the southwestern outskirts of Utrecht, some 60 km from the sea. A very busy motorway crosses over the canal near the plant and local industries include chemicals and food processing. Together all these various activities give rise to dust, salts, and fine aerosols in the air. The gas turbines ran on natural gas (no fouling of the hot section) and all power measurements were made with the gas turbine running in temperature control mode at base load hot gas inlet temperature. All results, see Fig. 1, have been corrected to new and clean guaranteed conditions.

The following was established during the above test.

- Without cleaning, the power output degradation tends to stabilize itself with increasing operating hours. It was confirmed in the unit tested, that the degradation of output was stabilizing at 90 percent base load (new and clean).
- Power recovery after off line washing (soak and rinse procedure) is significantly higher than after an on line washing.
- On line washings were performed at time intervals in the range between 700, 350, and 120 operating hours. It showed that plant performance is significantly higher at shorter on line washing intervals, thus preventing incremental power degradation.
- The combination of both washing methods is the most effective and economical.
- Based on the evaluation of the above performed measurements and extrapolated to 8000 operating hours, it was estimated that improved performance equivalent to approx. US $450,000 per year can be achieved with the combination of both on line and off line compressor washing methods on one 30 MW gas turbine, the operating and maintenance cost of the systems are not considered in the above figure.

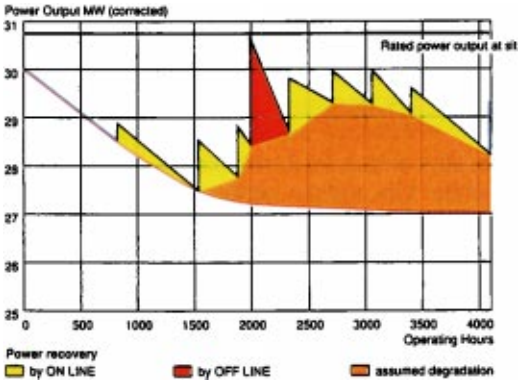


Fig. 1 Typical effect of on line and off line compressor wet cleaning

[1]This paper was awarded by IGTI the 1998 John P. Davis Award which recognizes gas turbine application papers which are judged to be of exceptional value to those supplying or using gas turbines and their support systems, further also for being an excellent contribution to the literature of gas turbine engine technology.

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Stockholm, Sweden, 1998; ASME Paper No. 98-GT-420. Manuscript received by the IGTI Division Mar. 1998; final revision received by the ASME Headquarters Nov. 2000. Associate Editor: R. Kielb.

Improvements With Shorter On Line Washing Intervals

In the Spring of 1990, UNA and Turbotect decided jointly to conduct a second long term field test on a 66 MW gas turbine operating in the Lage Weide 5 combined cycle plant located on the same site. The tests were conducted over 18 months under the combined on line and off line wet cleaning regime, from 18th May 1990 to 18th November 1991. During the entire test period the gas turbine unit operated for a total of 8089 h. An outage for a major overhaul (at 26,408 op. h) took place after 3915 operating hours since the beginning of the test. Thus the tests aimed to give some comparative indications of the effectiveness of more frequent on line washing as applied to a new machine, and to one that operated for several years, as well as on the plant's performance. The gas turbine also ran on natural gas. The air inlet filtration system consists of a weather louvre, a first stage coarse filter, and a second stage fine filter. For further references see Stalder and van Oosten [2].

Results of the Improved Compressor Cleaning Regime. On line compressor washes were performed, in the average every four days at base load with the gas turbine on temperature control mode. Gas turbine performance was measured before and after each wash.

Observations have shown a sustained high output level close to the nominal guaranteed rating, despite difficult atmospheric conditions.

In the first evaluation block, see Fig. 2, the gas turbine plant was operated at a load factor of 97.6 or 2.4 percent below the original guaranteed site power output at new and clean conditions. During this period 38 compressor on line washes were performed, in the average of one every four days. In addition, three off line washes were performed by taking the opportunity when the gas turbine plant was shut down for a few days, this, respectively, after intervals of 760, 2435, and 605 op h. The average power output increase after an off line wash was approximately 1800 kW. The trend analysis of the performance tests made in this period is nearly horizontal, showing that aging due to mechanical wear and tear of the gas turbine had already stabilized.

In the second evaluation block, Fig. 2, the gas turbine started in a practically new and clean condition as the result of some work made during the major overhaul. The corrected results of the compressor wet cleaning regime in the second evaluation block show that the gas turbine plant operated for 4174 h at a load factor of 100.16 or 0.16 percent above the original guaranteed site power output at new and clean conditions. At the end of this period the number of operating hours of the gas turbine was 30,725. During this second period, 45 compressor on line washes were performed, also on the average of one every four days. In addition, two off line washes were performed, one after 1143 and the second 1381 op. h later. The average power output increase after each off line wash in the second period was approximately 1 MW.

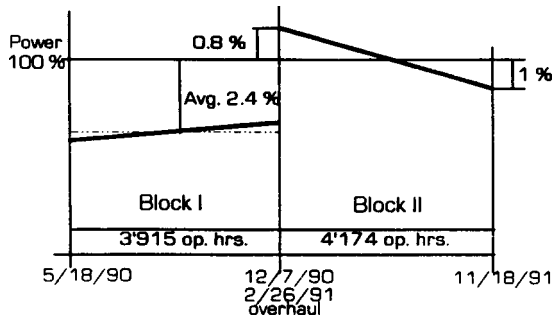


Fig. 2 Summarizes the pattern of the corrected power output for the complete test period

Discussion of the Results for the Improved Washing Regime.

- Out of the 83 on line washes made during the total testing period covering 8089 operating hours, 87 percent or 72 on line washes have demonstrated a positive power recovery with the unit in operation at full load.
- 712 kW was the average power output recovery measured after an on line wash. This relative small amount represents approx. 1 percent of the nominal gas turbine power output.
- No secured results on efficiency improvements could be demonstrated because of incomplete data over the testing period with regards to gas analysis and densities to determine the lower heating value, the latter being necessary to have accurate corrected efficiencies and turbine inlet temperature by heat balance calculations.
- Power recovery, due to off line cleaning, is not as significant, if the unit's performance is close to nominal guaranteed values.
- On line and off line cleaning are complementary.
- Shut-downs and start-ups can positively affect compressor fouling by spalling off deposits. The deposits may soak humidity during standstill and the swelled up material will partly spall off as the shaft is accelerated during start up of the gas turbine.
- The test program confirmed that frequent on line cleaning extends the time interval between off line cleaning operations. Thus it is a real benefit to the operator, because the scheduled down time allowed for maintenance can be reduced, if the frequency of off line cleaning with its associated cooling down time can be reduced. The availability and performance, as well as the overall profitability of the plant will be improved.
- The obtained results demonstrated that a combined on line and off line washing regime can effectively be applied to a new or to an old engine.
- The unit performance measured in the second evaluation period with above regime of washing shows that the power output trend was most probably following the aging of the unit.
- Also worth note is the impact of the major overhaul on the unit performance. The work involved in readjusting shaft alignment and clearances in the hot section path, etc., are likely to improve output and efficiency of the turbo group and can offset the cost for the overhaul.

Considerations on the Economical Profitability. The average power degradation over 4000 operating hours at the above tested plant and when compressor cleaning is not performed is up to 10 percent. By using similar criteria as presented by Diakunchak [3], the plant production profitability during above regime of on line and off line compressor wet cleaning was improved by US $1,175,000 over 8000 operating hours, representing a very substantial additional profit. An amount of approximately US $20,000 was spent for the chemical cleaner consumed during the program, representing a very marginal cost as compared to the improved profitability. Further, and without a regime of on line and off line compressor wet cleaning, there will be an additional loss of profitability due to reduced steam production as a result of compressor fouling in a combined cycle application. The reduction in mass flow has a greater effect on the steam production than the increase in exhaust temperature due to compressor fouling.

Compressor Fouling Phenomena

The cause of fouling and fouling rates of axial gas turbine compressors is a combination of various factors which can be divided into the following categories.

- Gas turbine design parameters.
- Site location and surrounding environment.
- Plant design and layout.
- Atmospheric parameters.
- Plant maintenance.

Gas Turbine Compressor Design Parameters. Smaller engines have a higher sensitivity to fouling than larger engines. Tarabrin et al. [4–19] concluded that the degree of the particles deposition on blades increases with growing angle of attack. Further, that the sensitivity to fouling also increases with increasing stage head. Multishaft engines are more sensitive to fouling than single shaft engines. Design parameters such as air inlet velocity at the inlet guide vanes (IGV), compressor pressure ratio, aerodynamical and geometrical characteristics will determine the inherent sensitivity to fouling for a specific compressor design.

Site Location and Surrounding Environment. The geographical area, the climatic condition and the geological plant location and its surrounding environment are major factors which are influencing compressor fouling. These areas can be classified in desert, tropical, rural, artic, off-shore, maritime, urban, or industrial site locations. The expected air borne contaminants (dust, aerosols) and their nature (salts, heavy metals, etc.), their concentration, their particle sizes and weight distribution, as well as the vegetation cycles and the seasonal impact are important parameters influencing the rate and type of deposition.

Plant Design and Layout. Predominant wind directions can dramatically affect the compressor fouling type and rates. Orientation and elevation of air inlet suctions must be considered together with the location of air/water cooling towers in a combined cycle plant, the possibility of exhaust gas recirculation into the air inlet, orientation of exhaust pipes from lube oil tank vapor extractors, as well as with other local and specific sources of contaminants such as location of highways, industries, seashores, etc.

Other plant design parameters which affect the rate of compressor fouling are as follows.

- The selection of the appropriate type of air inlet filtration system (self-cleaning, depth loading, cell, pocket, mat, pleat, oil-bath filter, etc.), the selection of filter media, the number of filtration stages, weather louvres inertial separators, mist separators, coalescer, snow hoods, etc. Design parameters such as air velocity through the filters, filter loading and their behavior under high humidity, pressure drops, etc., are also factors of high consideration.
- In case inlet conditioning systems are used, to have appropriate mist eliminators installed downstream of evaporative coolers. Inlet chilling in humid areas will result in continuous saturated conditions downstream. Thus, the presence of dust contamination in the air can combine with the moisture and additionally contribute to compressor fouling.

Plant Maintenance. Quality of air filtration system maintenance, frequency of compressor blade washing (deposition leads to higher surface blade roughness which in turn leads to faster rate of degradation), prevention of potential bearing seal oil leaks into air inlet stream, periodic water quality control in closed loop evaporative cooling systems, etc., are all measures which can positively influence compressor fouling and their rate of fouling.

Atmospheric Parameters. Ambient temperature and relative humidity (dry and wet bulb temperatures), wind force and direction, precipitation, fog, smog, or misty condition, atmospheric suspended dust concentration related to air density, air layer mixing by air masses, etc., are parameters which impact on the rates of fouling.

Correlation Between Rate of Power Degradation and Atmospheric Condition

Based on various field tests and observations, a correlation between rate of power degradation and atmospheric conditions prevailing at site can be established. As a result of the test conducted over 18 months under the combined on line and off line washing regime and presented earlier in this paper, we would like to first discuss herewith one of the most important observations made.

Disparity in Power Loss Gradients. Out of 40 measured continuous operating periods (without shut-downs and start-ups), a total of 14 operating periods each between 70 to 72 hours can be directly compared; the power output level at which the gas turbine was operated was always approx. the same. Power output measurements were made at the beginning of each period after on line washing (100 percent reference point) and at the end of each period, prior to on line washing of the next period. Figure 10 in Appendix A shows the power output losses over 70 operating hours. One can see that there is a widespread disparity in resulting power losses over such a short period, the highest loss in performance was 3.1 percent output and the end of one period shows even a gain of 0.5 percent power output. This surprising result was obtained on the same unit, with the same air inlet filtration system, the same washing nozzle system, the same washing procedure, and the same detergent.

Explanation for the Correlation Between Rate of Power Degradation and Atmospheric Condition. It is generally assumed that power losses will depend on the amount of humidity in a specific environment. With the data collected during the above comparative test periods, the total quantity of water and vapor mass flow ingested by the compressor was determined. The respective compressor air mass flows have been calculated by means of the heat balance. It was observed that the average ingested total humidity (water and vapor) amounted to 7.7 tons/h, or in total 548 tons during 70 operating hours. The lowest average value during a period was 4.1 tons/h and the highest was 11 tons/h. The plot below in Fig. 3 shows the measured power losses versus the total quantity of humidity (water and vapor) ingested by the compressor for each of the selected comparative operating periods of 70 h. see also Fig. 10 in Appendix A. The results of this test clearly indicates that there must be a correlation between the mass flow of absolute humidity and the loss in power output. Loss in power output increases with increasing mass flow of absolute humidity until it reaches, on the unit tested, a peak at approx. 400 to 450 tons (total over 70 h) before decreasing again.

Correlation Between Power Loss Gradients and Humidity. Due to the combination of pressure drop and increased velocity in the air inlet, humidity content in the air will start to condense in saturated condition. For instance and assuming a 250 MW gas turbine unit with an air velocity of 0.5 Mach at IGV's and an air mass flow of 500 m^3/s, operated at 12°C gas turbine compressor inlet temperature and 90 percent relative humidity (RH) as compared to 60 percent relative humidity, then the total condensing water mass forming droplets will be up to 6.3 tons/h at 90 percent RH as compared to 3.1 tons/h at 60 percent RH. The latent heat released by the condensing water will be higher at 90 percent RH, therefore the static temperature drop in the air inlet at 90 percent RH is 7°C whereas it is 10°C at 60 percent RH. See also the illustrated psychrometric chart in Fig. 4 below.

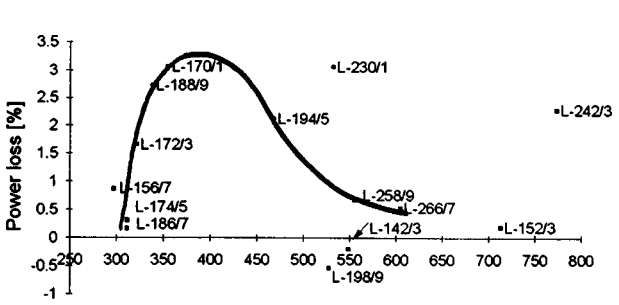


Fig. 3 Power losses vs. total absolute humidity for 14 comparative operating periods

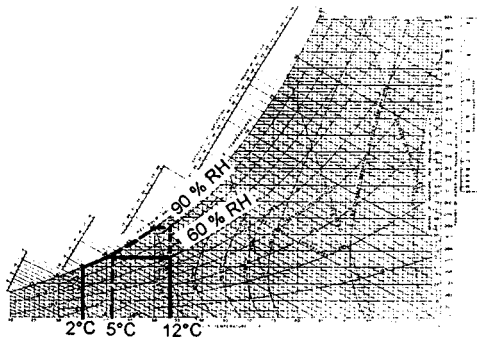


Fig. 4 Psychrometric chart for air inlet saturating condition

The above mass flow amount of condensed water droplets are impressive and give rise to the following explanation or combination thereof for power degradation patterns.

(a) **Influence of humidity.** Surface wetness of compressor blades operating in saturated condition will modify the aerodynamic boundary layer and cause a decrease in performance.

(b) **Latent heat release.** The ingested air temperature will increase at the compressor bellmouth entry as condensation occurs and latent heat is released, thus reducing cycle efficiency.

(c) **Water wettable and water soluble deposits.** With lower amounts of condensed water droplets, the ingested soil will combine with the water droplets and deposit on the vane and blade profiles. The rate of deposition will increase with the resulting roughness. Above a certain amount of formed condensed water droplet mass flow, the blading will be naturally washed and power losses due to fouling will be generally recovered to some extent.

(d) **In the case of hydrocarbon type of deposits.** The same effect as above will occur in the lower range of condensed water droplet mass flow. However, water being a poor hydrocarbon dissolver, the natural washing effect in the higher range of condensed water droplet mass flow will be very limited, if at all. Power losses due to fouling will continue until reaching equilibrium.

(e) **Combination of water wettable/soluble and hydrocarbon type of deposits.** This type of deposit is very common and pending upon the mass relationship of hydrocarbon vs. water wettable/soluble parts, and their respective embeddement in the deposition layers, the natural washing effect in the high range of condensed water droplet mass flow can also be very limited.

(f) **Fouling rate in the low range of ambient temperature.** The flattening of the saturation curve in the low ambient temperature range (<approx. −10°C), in the psychrometric chart, shows that the amount of water droplets which can combine with soil is very much reduced and thus lower fouling rates if at all, can be expected.

(g) **Duration and sequence of operation in saturated condition.** The changes in rate of power losses noticed over a given operating period will be very much influenced by the duration and sequences of operation in saturated condition (vs. dryer condition). For instance, it is most probable that a high amount of water droplet mass flow ingested in the beginning of an operating period with a clean compressor will affect the changes in rate of power loss differently then if it would occur towards the end of such an operating period.

Discussion on Washing Technology, Acquired Know How and Experiences

Fouling Deposits. Most fouling deposits are mixtures of water wettable, water soluble, and water insoluble materials. Very often pH 4 and lower can be measured in compressor blade deposits. This represents a risk of pitting corrosion. Furthermore, these deposits become more difficult to remove if left untreated as the aging process bonds them more firmly to the airfoil surface, thus reducing cleaning efficiency.

Water soluble compounds cause corrosion since they are hygroscopic and/or contain chlorides that promote pitting corrosion. They can be rinsed, however, they can also be embedded in water insoluble compounds. Water insoluble compounds are mostly organic such as hydrocarbon residues or from silica (Si).

On Line and Off Line Compressor Cleaners. The cleaners available today in the market are generally nonionic and designed to fulfill gas turbine engine manufacturers specification for both on line and off line cleaning, thus also simplifying stock keeping and handling on site. Off line cleaning being done at crank speed, crank-soak washing method, with the engine cooled down (requires cooling down time), whereas on line cleaning is made during operation of the engine. The used detergent and demineralized water in on line cleaning application must generally fulfill the manufacturers fuel specification.

Function of a Compressor Chemical Cleaner. The main constituent of a cleaner is its surfactant (surface active agent) the purpose of which is to reduce surface tension of the solution enabling it to wet, penetrate, and disperse the deposits. Such rapid intimate mixing cannot be achieved with water alone. Surfactants are therefore needed for water insoluble deposits—both liquid and particulate types—to enable their removal from compressor blade surfaces and to prevent redeposition.

Foaming of the Compressor Cleaner. The amount of foam generated by a compressor cleaner is an indication of the degree of activity and therefore of the effectiveness of the surfactants used in the cleaner. Foam being light, will help to achieve a better distribution and penetration of the cleaning solution into the deposits during off line washing, it will keep moisture inwards and thus extend the contact time. The foam will be rapidly displaced during rinsing and will help to remove surfactants left on the blade surface. Water films alone will tend to drain off the blades more rapidly thus reducing contact time during the off line soaking period. In addition, foam also acts as a dirt carrier.

Corrosion Inhibitors/Compressor Rinse. Some compressor cleaners do contain a corrosion inhibitor claimed to inhibit corrosion by neutralizing the influence of salts during a certain period and thus sparing a time-consuming dryout run after off line washing. This may be important for jet engines that are frequently not reoperated immediately after compressor crank washing. However, the above situation is different in a stationary gas turbine generating plant. pH or conductivity measurements in off line effluent water have shown that salts are often left behind to some extent after washing with the cleaner agent only, thus rinsing with water alone will be beneficial to remove the salt solution after washing, consequently a blow-run followed by a dry-out run is recommended. The blow-run at crank speed shall help to drain all trapped water in the GT internal piping systems and the completed dry-run at no-load shall give the assurance that the mechanical condition of the GT has been reestablished and that the latter is ready for start by the plant dispatcher.

Corrosion inhibitors have a high affinity to surfaces and will tend to form a film on the blades, therefore there is a high potential risk during on line cleaning that this film can produce a decomposed material deposit in the temperature range of 200°C, in the middle section of the compressor. Our recommendation is to eliminate the salts left behind by doing a demineralized water rinse of the same duration as the cleaning period with detergent after each on line cleaning.

Hot or Cold Water for Off Line Washing. Hot wash water will soften the deposits better than cold water, further it will also help to prevent thermal shocks and thus reduce the cooling time period. However, hot water has also several disadvantages, it has

to be hot enough (in the range of 60 to 80°C), which means that the wash skid will require a heating system, heating energy, an insulated tank, and piping. The heater will have to be started approx. 12 h before washing (depending on heater capacity). Thus, washing will have to be scheduled in advance. Whereas many users are washing their engine off line by opportunity. Using cold water will reduce the manufacturing and installation cost of the washing skid as well as the aux. energy consumption (they can be significant for large engines having a high cleaning solution injection mass flow). Using cold water (approx. >15°C) will also significantly increase maintenance scheduling flexibility. A weaker softening of deposits by cold water is by far compensated by more frequent washing (combined on line and off line washing regime) and good cleaner surfactant performance.

Water Based Cleaners Against Solvent Based Cleaners. Water based cleaners have a broader range of application as they are able to deal with oily soils, water soluble salts and particulates (silica, clay, etc.). Solvent based cleaners have a more specific application where the oily portion in deposits is predominant. For off line washing, to facilitate effluent water disposal, water based cleaners must be easily bioeliminable and biodegradable and they should comply to the Organization of Economic Cooperation and Development (OECD-302B) requirements. Whereas, when using solvent based cleaners, effluent water generally needs to be handled and treated as used oil. Cleaner selection is normally dictated by the type of deposits, the available effluent water disposal system and the relevant local regulation for effluent water disposal.

For on line washing, the selection and the quality of the type of cleaner can play a very important role. Water based cleaners diluted with demineralized water in ratios as recommended by detergent manufacturers have much lower residual formation (normally <0.1 to 0.7 percent at 200°C, see Fig. 5, whereas solvent based cleaner can have up to 3 percent or more of residual formation at 200°C. Turbotect also recognized in a very early stage that the selection of raw materials used in the manufacture of the cleaners for both solvent and water based cleaners is most important. An inappropriate selection of raw materials can lead to the formation of resinlike deposits or oily films on blade profiles during the decomposition, generally in the middle section of the compressor. Such sticky deposits can collect dirt washed off from the front and foul the compressor downstream. Thus, frequent washing may lead to an undesired result whereby no power recovery after washing will be achieved. Properly selected raw materials in the manufacture of cleaners will not lead to the formation of such gummy build-up deposits. In order to reduce likewise deposits downstream of the compressor, it is also preferable to use water based cleaners for on line washing (lower residual content).

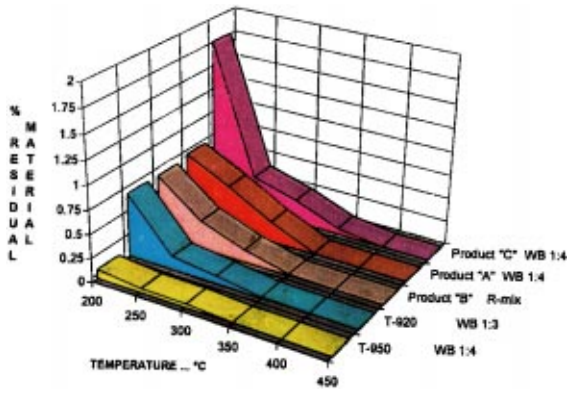


Fig. 5 Material residual percent vs. temperature for water based product. Note that product "C" showed an oil film at 200°C. (WB: Water based cleaner; R-mix: Ready for use WB cleaner.)

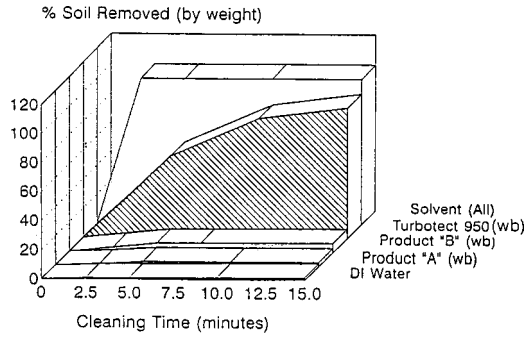


Fig. 6 Comparison of cleaning efficiency. (WB: Water based cleaner; DI-water: Demineralized water.)

Solvent based cleaners do have a better cleaning efficiency against oily type deposits. Figure 6 below shows cleaning efficiencies on oily deposits achieved by demineralized water and various water based and solvent based cleaners. Note that cleaning efficiency will also be affected by the injection time.

Compressor On Line Cleaning With Water Only. Some plant operators are doing on line washing regime tests by comparing the power recovery results using demineralized water only against using a detergent cleaner. The value of such tests can be very important because it will help the plant users to understand the phenomena of compressor fouling by monitoring very closely the unit's performance and to experiment various washing regimes. As already mentioned before, compressor fouling behavior is plant specific. Water as on line cleaning media will certainly work, however, only if the deposits are totally water wettable and/or water soluble. We believe that a minority of engines will have only such deposits. Using water only as cleaning media will impose very short time intervals between washings to avoid any build-up at all. As otherwise it can be detrimental in case the deposits are water insoluble or a combination of water soluble and insoluble compounds, the insoluble part will not be washed off, thus allowing further build-up of deposits in the front stages. Therefore, we generally recommend the use of water based cleaners for on line washing, which leads us to the subject of optimum washing regime.

Detergent Use and Optimum Washing Regime. An optimum washing regime shall keep power decay to a minimum by applying combined on line and off line cleanings. An on line washing period shall always be started with a clean engine after an overhaul or an off line cleaning. The time intervals between on line washing shall be kept short, approx every 3 days to every day. However, depending on the type of deposits (portion of insoluble compounds), detergent cleaners may be used for every second or third on line wash or up to once a week only. Between times, on line cleaning can be made with demineralized water only. In case the time intervals between detergent on line washes is extended, there will be a higher risk of downstream contamination due to larger portions of insoluble compounds being lifted at once and carried through the compressor. Thus, we recommend frequent on line cleanings with detergent in order to have only minute portion of insoluble compounds being washed off at any given time.

On Line Compressor Cleaning Efficiency. Deposits on the profile of the first stage vanes are primarily responsible for a significant reduction in air mass flow through the compressor, thus reducing the power output. On line cleaning is most effective in removing such deposits on the first stage guide vanes, thus restoring design air mass flow, i.e., see Figs. 7 and 8. Frequent on line cleaning keeps the first stage guide vanes clean. Droplets of cleaning solution may survive up to the sixth stage, however most are vaporizing by then and the residue/ashes will be centrifuged along

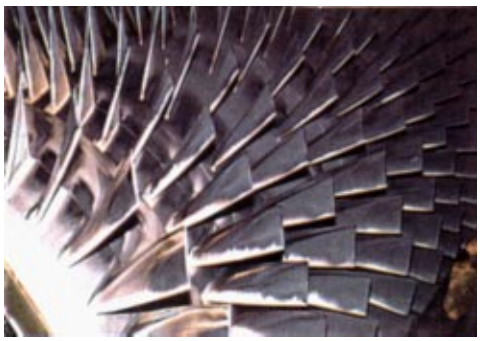


Fig. 7 Picture of a compressor which was washed on line every four days with detergent during approx. a one month continous operation period

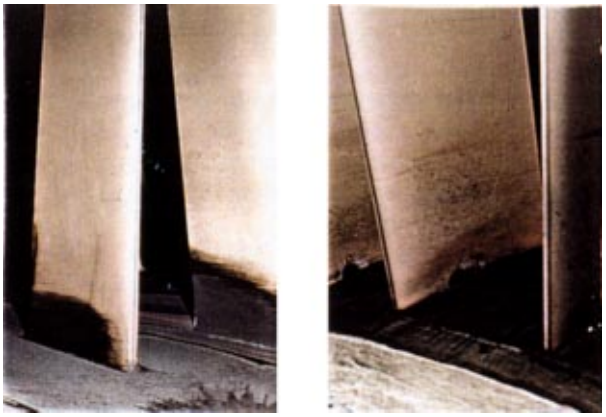


Fig. 8 First stage guide vanes of the same compressor (pictured in 1/3 vane length from the tip and from the root), washed on line every four days with detergent during approx. a one month continous operation period

the compressor casing. Therefore, no cleaning solution will be available for removing deposits on downstream stages.

The above pictures of the IGV (Fig. 8) are very interesting and show a very effective on line washing, when observing the film deposit on the blade tip, one can easily conceive the severeness of power loss and deposits on this first stage if no on line wash had been applied by the operator.

Off Line Compressor Cleaning Efficiency. The off line cleaning method is very efficient for removing all deposits on all the compressor stages. For this purpose off line washing shall preferably be performed at variable crank speed, e.g., by injecting the cleaning and rinsing solution during coast down of the shaft after an acceleration of up to say 500 to 600 rpm. By doing this, the pattern of the centrifugal forces on the injected solution through the compressor will decrease and allow a better wetting and distribution on the blade and vane surfaces of all stages. By contrast, off line washing at high and constant cranking speed will end up in lower off line cleaning efficiency. Conductivity measurements in rinsing water as well as checks on the clarity/turbidity of rinsing water will help to assess the cleaning efficiency.

Off Line Cleaning Intervals. Irrespective of the compressor performance degradation, a sound combined on line and off line washing regime should incorporate at least four off line compressor cleanings per year in order to remove the salt laden deposits on the downstream stages. This recommendation being also valid for a peak load unit running only few hours per year.

Washing Equipment and Procedure

System Engineering and Equipment. The following two basic principles have been determined by Turbotect for the engineering of both on line and off line injection systems.

- Limit the injection mass flow to what is absolutely required for a good cleaning efficiency.
- Totally reliable nozzle injection system, giving due consideration to all operational aspects.

On Line Cleaning Injection System Design Criteria. The objective is to achieve the highest washing efficiency at the smallest injection mass flow for the following essential reasons.

- **On line injection mass flow.** Gas turbine engine manufacturers are very much concerned with deposition of washed off dirt from the front stages onto downstream compressor stages. Further, that washed off dirt may enter into sophisticated airfoil film cooling systems of turbine blades with the potential to clog them, resulting in temperature hot spots. In this respect, also a sound isolation scheme to avoid run-off effluent water penetrating into the internal engine piping systems during off line washing is of prime importance. Low level of flame detector intensities during on line cleaning on units with Dry Low NO_x (DLN) combustors are also claimed for potential tripping of the units (e.g., fogging of flame detector lenses, etc.). Some users also observe higher CO emission levels during on line washing. Small quantity on line cleaning injection mass flow combined with the optimum washing regime will counteract against above claimed potential problems in modern gas turbines. Therefore, a small injected quantity of on line cleaning solution mass flow is preferable. For instance, Turbotect's on line compressor cleaning system for a 120 MW heavy duty engine comprises 30 injection nozzles, the total mass flow applied is only a fraction of comparative on line washing systems. This low mass flow has demonstrated that it does not impair the cleaning efficiency and has further also the advantage to lower significantly overall demineralized water and cleaner consumption. In addition it will also reduce the required size, volume, and cost of the washing skids.
- **Wetting.** A very effective wetting of the IGV's is reached by a uniform and finely distributed atomized cleaning solution. Droplets are subject to gravity, they must be stable in size and small enough that they do not cause blade erosion, and light enough that they do not drop out of the air stream before they reach the compressor blade surface. A nonuniform wetting of the IGV's will result in spot cleanings and heavier droplets will most likely fall to the bottom, wasting some injected cleaning solution.
- **Equipment**. The on line nozzle injection design is of prime importance to achieve a high washing efficiency. The nozzles are designed to inject a small quantity of fine atomized cleaning solution into the air stream where it will be thoroughly mixed and carried uniformly into the compressor bellmouth. A relatively high number of nozzles positioned in the air inlet casing on both up-stream and down-stream sides of the bellmouth are ensuring a better distribution of the injected fluid into the air stream and consequently a better wetting.

The nozzle atomizer is integrated in a spherical body which can be rotated in two dimensions to set the spray angle. The adjustment of the spray angle allows proportioning of atomized cleaning solution massflow which shall penetrate into the air stream from the mass flow which shall remain trapped in the boundary layer for root and tip blade cleaning. The nozzle body is installed flush mounted under the surface of the intake structure and penetration of the nozzle into the air stream is minimal to avoid inducing vibration into the air stream. Further, this design also prevents misuse of the nozzles as climbing support during compressor intake inspection. Its design is ruling out loss of any parts into the air stream (see Fig. 9).

Off Line Cleaning Injection System Design Criteria. The

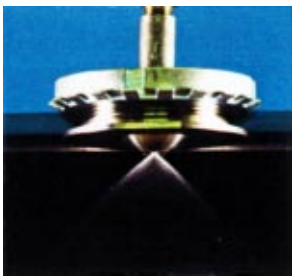


Fig. 9 On line injection nozzle (patented) allowing orientation and fine tuning of the spray in any direction

objective again is to achieve the highest washing efficiency at the smallest injection mass flow for the following essential reasons.

• **Off line injection mass flow.** First at all it is obvious that users are interested in low quantity effluent water to be disposed off. It is also claimed that off line water effluent transported up to the exhaust during off line washing may wet and soak into the expansion joint fabric (i.e., or also in recuperator compensator). Thereupon some expansion joint fabric may lose some of their thermal insulation capacity, resulting in overheating of the fabric which can damage the expansion joint.

• A low off line injection mass flow will also reduce the potential risk and measures to be taken against trace metals and alkali compound contaminations in exhaust systems where selective catalytic reactors (SCR) for NO_x reduction or CO catalysts are installed. A low off line injection mass flow will significantly reduce the required size, volume and cost of washing skids and consequently the overall water and cleaner consumption.

• **Wetting/Equipment.** A very effective wetting of the IGV's suction area is achieved by using full cone jet spray nozzles. The number of nozzles will be defined by the area to be wetted, usually the area between two struts. The necessary off line injection mass flow characteristic will therefore be determined by the area to be wetted and impacted by the jet spray and the distance between their location of installation in the inlet casing up to the area of impact, the injection pressure is generally between 5.5 to 6 bars. The spray jets are also subject to gravity, the nozzle is designed such that an angle (up to 5 deg) can be adjusted for compensation. The off line washing (soaking) and rinsing method can be considered as a mechanical erosion of deposit layer and soaking time will allow the cleaner to penetrate and soften the deposit layers. For instance, systems with high atomization pressure will have no impact pressure on the IGV's, this because the spray pressure will have lapsed approx. 20 cm from the nozzle outlet and the atomized droplets will need to be carried by the small air stream produced at crank speed. These high pressure systems do not show the same effectiveness in removing salts and insoluble compounds on downstream stages.

• **Drainage.** The effluent water collection system to drain the dirty water outside the engine and the isolation scheme to hinder run-off water to penetrate sensitive systems such as sealing and cooling air systems, instrumentation air systems, etc., during washing is of prime importance.

The physical location of the air systems taps on compressor casing is also important. Taps on the bottom are likely to drain run-off water, they should be preferably located in the upper part of the casing. Drains in air inlet casing, in the compressor casing, combustion chamber, and exhaust should be located at the lowest point. The drain diameters should allow a good run-off, it can be observed that sometimes they get plugged because of dirt being not properly evacuated. A good isolation and drainage concept will insure that no dirty water ingress into cooling, sealing-air, and other systems during the off line washing procedure.

Paint and Corrosion Damage in Air Inlet Manifolds. Some users are claiming that the corrosion protective paint applied originally inside the air inlet casing gets softening and some detaching paint spots are depositing by the air stream on IGV's due to frequent on line washings, this possibly because of the chemistry involved in the cleaner.

Chemistries involved in the formulation of compressor cleaners are very soft. These cleaners have all been tested for e.g. immersion corrosion, sandwich corrosion, and hot corrosion on metals, alloys, coatings, etc., used in the manufacture of compressors; effect on polymeric materials, on rubber, silicone elastomers, on epoxy adhesives, on painted surfaces, etc., according to existing US MIL-C-85704B specification ''Cleaning Compound, Turbine Engine Gas Path.''

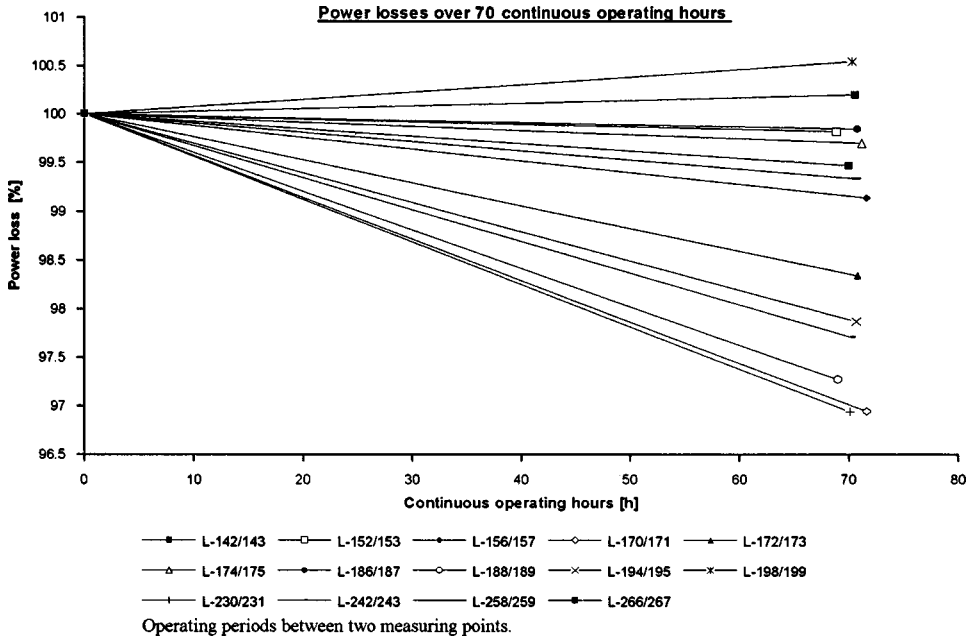


Fig. 10 Appendix A

Investigations conducted together with paint manufacturers have shown that softening and detaching paint systems in air inlets are caused for various reasons or combination thereof; even paint systems have shown damages in units where no on line washing were performed. Casings are also very often wetted by the existing condensed water generated in saturated condition. Herewith some causes for potential paint degradation in air inlet casings: Quality of surface preparation was not sufficiently carried out (sharp edged material to rough the surface by blasting is a prerequisite). Recommended paint thickness, drying time and application temperatures to be followed. In case of paint repairs, old paint should be grinded away and not removed with liquid paint removers. Combination of selected paint systems (type, 1 or 2 component systems, etc.) for priming, first coat and finish coating are not compatible. Some paint systems show swelling characteristics.

Outlook for the Future

Close monitoring of compressor performance, atmospheric parameters, together with performance evaluation and histograms will help improve washing regimes, and possibly to predict fouling and energize automatically on line washing whenever necessary. Such aptitudes will definitely improve overall plant profitability together with well engineered compressor washing systems. Many users are also facing severe fouling at ambient temperatures below $+10°C$ where compressors operate under conditions where icing becomes a risk. On line washing in this temperature range and below is the next coming challenge to be addressed.

Conclusions

Fouling rates can vary very much and are very specific to each plant. The gas turbine unit design parameters, the site location and its surrounding environment, the climatic conditions, the plant concept, design, and its layout are given once the plant has been built. The site weather parameters are then having the largest impact on fouling rates and performance degradation. Therefore improvements on the plant profitability by reducing the impact of compressor fouling can only be related to the plant maintenance, monitoring of plant performance, the performance of the installed washing system, the selection and quality of the detergent used and the optimum chosen washing regime.

References

[1] Hoeft, R. F., 1993, ''Heavy Duty Gas Turbine Operating and Maintenance Considerations,'' GER-3620B, GE I&PS.

[2] Stalder, J. P., and van Oosten, P., 1994, ''Compressor Washing Maintains Plant Performance and Reduces Cost of Energy Production,'' ASME Paper No. 94-GT-436.

[3] Diakunchak, I. S., 1991, ''Performance Deterioration in Industrial Gas Turbines,'' ASME Paper No. 91-GT-228.

[4] Tarabrin, A. P., Schurovsky, V. A., Bodrov, A. I., and Stalder, J. P., ''An Analysis of Axial Compressor Fouling and a Cleaning Method of their Blading,'' ASME Paper No. 96-GT-363.

[5] Aker, G. F., and Saravanamuttoo, H. I. H., 1988, ''Predicting Gas Turbine Performance Deterioration Due to Compressor Fouling Using Computer Simulation Techniques,'' ASME Paper No. 88-GT-206.

[6] Bird, J., and Grabe, W., 1991, ''Humidity Effects on Gas Turbine Performance,'' ASME Paper No. 91-GT-329.

[7] Haub, G. L., and Hauhe, W. E., Jr., 1990, ''Field Evaluation of On Line Compressor Cleaning in Heavy Duty Industrial Gas Turbines,'' ASME Paper No. 90-GT-107.

[8] Jeffs, E., 1992, ''Compressor Washing On Line for Large Gas Turbine,'' Turbomachinery International, **33**, (6).

[9] Kolkman, H. J., 1992, ''Performance of Gas Turbine Compressor Cleaners,'' ASME Paper No. 92-GT-360.

[10] Kovacs, P., and Stoff, H., 1985, ''Icing of Gas Turbine Compressors and Ways of Achieving Uninterrupted Operation,'' Brown BoveriRev. **72**, pp 172–177.

[11] Lakshminarasimha, A. N., Boyce, M. P., and Meher-Homji, C. B., 1992, ''Modelling and Analysis of Gas Turbine Performance Deterioration,'' ASME Paper No. 92-GT-395.

[12] Mezheritsky, A. D., and Sudarev, A. V., 1990, ''The Mechanism of Fouling and the Cleaning Technique in Application to Flow Parts of the Power Generation Plant Compressors,'' ASME Paper No. 90-GT-103.

[13] Saravanamuttoo, H. I. H., and Lakshminarasimha, A. N., 1985, ''A Preliminary Assessment of Compressor Fouling,'' ASME Paper No. 85-GT-153.

[14] Sedigh, F., and Saravanamuttoo, H. I. H., 1990, ''A Proposed Method for Assessing the Susceptibility of Axial Compressors to Fouling,'' ASME Paper No. 90-GT-348.

[15] Stalder, J. P., 1992, ''Professional System Approach to Compressor Cleaning, Case Studies,'' GCC-CIGRE Paper, Third Annual Conference, Dubai, May 1992.

[16] Thames, J. J., Stegmaier, J. W., and Ford, J. J., Jr., 1989, ''On Line Compressor Washing Practices and Benefits,'' ASME Paper No. 89-GT-91.

[17] Ul Haq, I., and Saravanamuttoo, H. I. H., 1993, ''Axial Compressor Fouling Evaluation at High Speed Settings Using an Aerothermodynamic Model,'' ASME Paper No. 93-GT-407.

[18] Ul Haq, I., and Saravanamuttoo, H. I. H., 1991, ''Detection of Axial Compressor Fouling in High Ambient Temperature Conditions,'' ASME Paper No. 91-GT-67.

[19] Zaba, T., and Lombardi, P., 1984, ''Experience in the Operation of Air Filters in Gas Turbine Installations,'' ASME Paper No. 84-GT-39.

A Neural Network Simulator of a Gas Turbine With a Waste Heat Recovery Section

C. Boccaletti

G. Cerri

B. Seyedan

Department of Mechanical and Industrial Engineering, University ROMA TRE, Roma, Italy

The objective of the paper is to assess the feasibility of the neural network (NN) approach in power plant process evaluations. A ''feed-forward'' technique with a back propagation algorithm was applied to a gas turbine equipped with waste heat boiler and water heater. Data from physical or empirical simulators of plant components were used to train such a NN model. Results obtained using a conventional computing technique are compared with those of the direct method based on a NN approach. The NN simulator was able to perform calculations in a really short computing time with a high degree of accuracy, predicting various steady-state operating conditions on the basis of inputs that can be easily obtained with existing plant instrumentation. The optimization of NN parameters like number of hidden neurons, training sample size, and learning rate is discussed in the paper. [DOI: 10.1115/1.1361062]

Introduction

Usually, industrial cogeneration plants are composed by sections equipped with similar parallel units and produce power and heat products like hot water, superheated steam, and so on. To generate these products, and to satisfy the facility demands, fuel enters the cogeneration plant and electricity is usually sold to (or bought from) the external National (Regional) grid. Planning the plant production capacity and the load allocation in the various sections and units requires the optimization of a cost function which takes the various aspects (plant initial cost, maintenance, consumables, and emissions of pollutants) into consideration.

During the last few years industrial plants have been equipped with electronic control and supervision systems which receive signals coming from the various measurement stations (pressures, temperatures, mass flow rates, and so on) distributed in the plant. Such data may be used to recognize the actual status of components by establishing status parameters to be imported into plant simulators.

The optimization requires iterative calculations which may take a long computing time when physical and empirical simulators are used. Equations are usually nonlinear and this introduces some difficulties of solution, which are also connected with the really high accuracy levels in predicting the plant behavior that are needed for cost optimization. Thus, the on-line use of optimization tools requires the adoption of numerical simulators performing only straightforward calculations.

The aim of this paper is to demonstrate that neural network (NN) techniques may be used to establish accurate numerical simulators of the plant units.

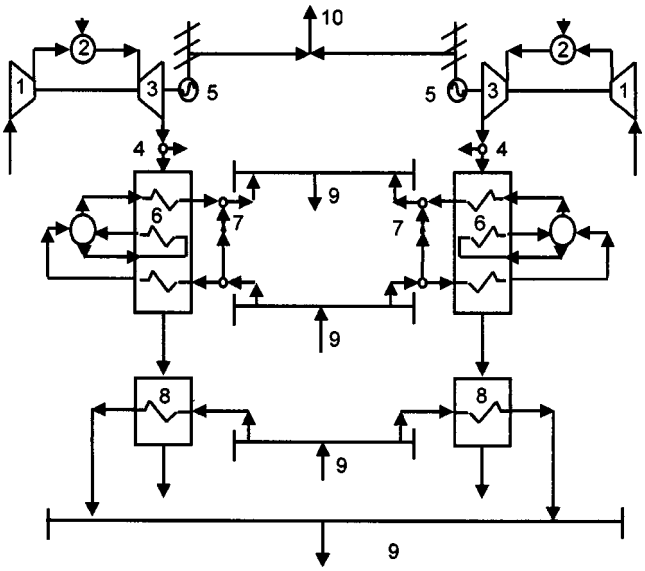
A cogeneration plant unit has been selected ([1]) that consists of a single shaft heavy duty gas turbine driving an electric generator. The exhaust gases from the gas turbine pass through a single pressure level steam generator and a water heater.

The physical simulator, developed for the optimum load allocation ([2]), has been used to produce the data base for NN training and testing.

After a brief description of the features and peculiarities of the unit and of the physical simulator, the development of the NN simulator is discussed together with the results.

Description of the Unit

A schematic diagram of the investigated industrial combined gas-steam section of a cogeneration plant is shown in Fig. 1. The section consists of two gas turbines (parallel units) whose exhaust gas available heat is recovered in the waste heat boilers (consisting of economizer, evaporator, and superheater) and water heaters. Electricity generated in gas turbines and steam and water from waste heat recovery devices are supplied according to the different facility demands. There is a possibility of bypassing exhaust gases to the stack (Fig. 1, station 4). Steam and hot water



1= compressor; 2= combustion chamber; 3= gas expander;
4= gas bypass; 5= a.c. generator; 6= waste heat boiler;
7= valve and de-superheating; 8= water heater;
9= water/steam manifolds; 10= electric grid.

Fig. 1 Cogeneration plant

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Munich, Germany, May 8–11, 2000; Paper 00-GT-185. Manuscript received by IGTI February 2000; final revision received by ASME Headquarters January 2001. Associate Editor: M. Magnolet.

temperatures are controlled by an attemperator and by water mass flow rate adjusting. The numerical prototype is related to a turbogas (TG) unit.

The Physical Simulator

A modular numerical prototype has been designed and programmed. The computer models are generic in nature and apply to any manufacturer's components. Each module is described by conservation equations at the steady state (mass, energy, and momentum), by the second law of thermodynamics, by constitutive and auxiliary equations describing phenomena on physical or empirical basis, and by control equations.

Modules that have been developed are compressor, combustion chamber, expander, heat exchange devices (economizer, evaporator, superheater), pumps, valves, connections, splitters, junctions, electric generator, and electric grid. Variables entering the single module are quantities related to the thermodynamic state and to geometric and global parameters that are suitable for representing the functional characteristics of the component and its specifications. The model of a plant component is adjusted according to the kind of problem to be solved, design (inverse) or part load (direct) analysis. The solution of inverse and direct problems for the heat exchangers is achieved through an ε-NTU (effectiveness–number of transfer units) method. Turbomachine behavior is described by a one-dimensional approach accounting for quantities like angles of incidence and deviation, loss coefficients, mass flow rates, pressure, temperatures, and so on. Losses and flow deviations are estimated by empirical correlations. Detailed descriptions of the developed component modules have already been presented in previous papers ([2,3]).

The component simulation implies the prediction of the desired output by solving a problem of constrained optimization. The solution is achieved once an appropriate unbalance function has vanished, satisfying a partial objective function. Cerri [3] has already described the solution method in detail. This method has been implemented to solve several problems described by Bohn et al. [4,5] and Cerri et al. [2,3,6,7].

In this case the minimum differences between the desired and calculated superheater outlet steam temperature and the desired and calculated hot water temperature have been obtained.

The unit boundary conditions are

1 ambient pressure, temperature, and relative humidity;
2 electric power;
3 actual statuses of components;
4 bypass gas flow rate.

Operating conditions have been predicted as a function of the above quantities. Calculated variables are fuel mass flow rate, combustion chamber exit temperature, blade and gas temperatures, compressed air and exhaust gas temperatures, steam and water mass flow rates, and so on.

A very accurate simulation has been performed, since small errors in the expected behavior of the components may cause great differences from the point of view of optimization. The simulator modules have been calibrated with experimental data. Figure 2 shows the gas turbine exhaust temperature versus power output. In this figure the solid line represents the expected curve, the circles represent the in situ measurement points, and the triangles represent the points calculated by the simulator. It may be observed that the simulator is able to give an accurate response in the entire range. A map representing the TG unit behavior is given in Fig. 3, in the case of exhaust gas bypass completely closed and hot water temperature at the exit of heater fixed at 140°C. The figure shows the variation of gas turbine exhaust temperature, steam temperature, hot water, and steam and fuel mass flow rates as a function of the produced electric power. The calculated fuel mass flow rates are compared with the measured ones. The agreement is good. It may be seen that the steam temperature rises with the gas turbine power, and then stabilizes due to the attemperator intervention. Steam and hot water mass flow rates increase approximately linearly with gas turbine power.

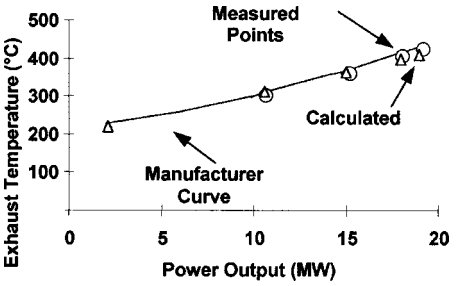


Fig. 2 Gas turbine power output versus gas expander exit temperature

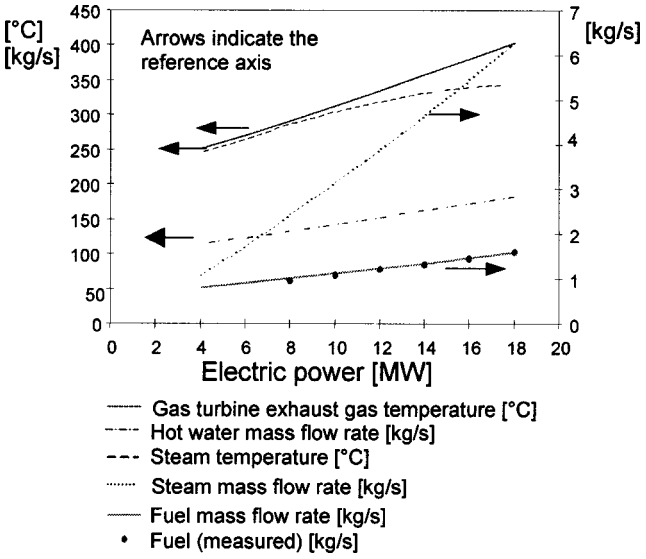


Fig. 3 Waste heat recovery and water heater section

Background of the Neural Network Approach

NNs have been applied successfully to a number of engineering problems, which are highly nonlinear in nature. In recent years renewed interest is due to the development of new net topologies and algorithms ([8–10]). NNs can represent the physical knowledge of complex systems. Physical rules contain the physical knowledge of the system behavior. NNs can extract such knowledge from the data presented to them and reproduce it.

A NN derives its computing power from massively distributed structures and from its ability to learn from examples. NNs may learn complex relationships among plant parameters from available experimental or simulated data. Neural computing is fault tolerant because the knowledge is not contained in one place, but is distributed throughout the system. If some processing elements are destroyed the behavior of the network as a whole is only slightly altered. If more processing elements are destroyed the behavior of the network will degrade just a bit further, whereas traditional computing systems are rendered useless by even small amounts of damage to the memory.

Figure 4 shows the scheme of an artificial neuron. Input quantities x_i are linearly combined using the weights w_i, then multiplied by an activation function to obtain the output. Different types of activation function can be used: sigmoidal, linear, hyperbolic tangent, logistic, etc. For the present study the logistic type activation function has been chosen.

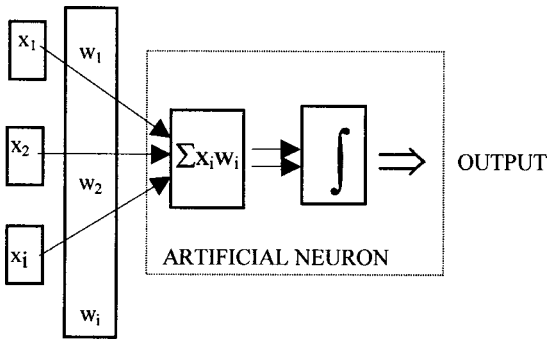


Fig. 4 Artificial neuron with activation function

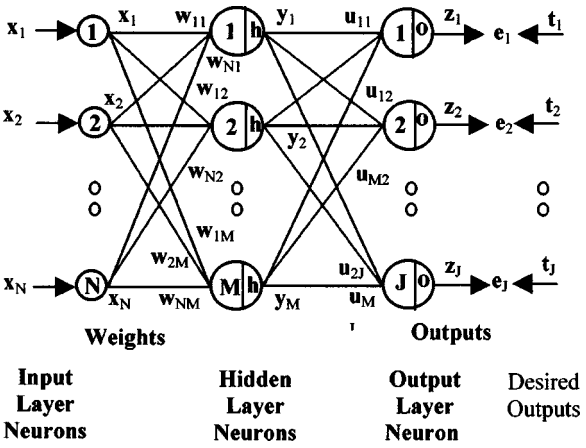


Fig. 5 Feed-forward artificial neural network

Single artificial neurons are combined to build a network. Various architectures are possible. The literature and the authors' previous experience ([11,12]) suggest that feed-forward NNs with a back propagation algorithm are the most suitable for the subject of this work.

A schematic architecture of the feed-forward NN considered for the present study is shown in Fig. 5. The first layer on the left contains the input neurons, the middle layer contains the hidden neurons, and the layer on the right the output neurons. Although architectures with multiple hidden layers can be applied, it has been shown that one hidden layer is usually sufficient. The input vector $\mathbf{x}=(x_1, \ldots, x_N)$ is linearly combined by the weights $\mathbf{w}=(w_{11}, \ldots, w_{NM})$ to obtain the output $\mathbf{y}$. Vector $\mathbf{y}=(y_1, \ldots, y_M)$ and weights $\mathbf{u}=(u_1, \ldots, u_{MJ})$ bring to the neural network output $\mathbf{z}_i=\Sigma u_{mj} y_m$. Vector $\mathbf{t}=(t_1, \ldots, t_J)$ is the desired output, and vector $\mathbf{e}=(e_1, \ldots, e_J)$ is the error, calculated as the difference between the output vector $\mathbf{z}$ and the target $\mathbf{t}$.

Topology, node characteristics, and training or learning rules characterize the NN models. The latter specify an initial set of weights and indicate how these weights should be updated to improve performance.

Setup of the Neural Network Simulator

The aim of the research activity was to produce a simulator reproducing the behavior of the turbo-gas unit (consisting of gas turbine, waste heat boiler, and water heater) using the NN approach. Input values have been chosen according to the application within physically acceptable ranges. Such variations of input and output quantities are shown in Table 1. The database has been created by means of the physical simulator. In order to train the NN, input and output quantities have been divided into two parts:

Table 1 Ranges of variability

#	Quantity	Unit	Range	
			Min.	Max.
Simulator input data				
1	Ambient pressure	kPa	90	102
2	Ambient temperature	°C	4	40
3	Relative humidity	%	0	100
4	Gas bypass	%	0	99
5	Electric power	MW	7	20
Simulator calculated values				
1	fuel mass flow rate	kg/s	0.8	1.7
2	air mass flow rate	kg/s	95	130
3	gas mass flow rate	kg/s	96	131
4	compressor exit pressure	kPa	561	748
5	compressor exit temp.	K	521	578
6	combustion chamber exit pressure	kPa	539	728
7	combustion chamber exit temperature	K	825	1110
8	blade temperature	K	825	1110
9	expander exit pressure	kPa	90	104
10	expander exit temperature	K	542	744
11	waste heat boiler exit press.	kPa	89	103
12	waste heat boiler exit temp.	K	478	581
13	stack temperature	K	393	446
14	steam mass flow rate	kg/s	0	8
15	steam temperature	K	541	614
16	superheated steam mass flow rate	kg/s	1.2	189

Table 2 Sample size distribution

Data Sample size Turbo-Gas Unit	
Training	4000
Validation	964
Verification	1679

training and verification data sets (see Table 2). The validation (test) data set consists of a subset of the training data.

Training. NN training is a theoretical issue that has attracted great attention in the literature and continues to do so. NN behavior is influenced by the size and efficiency of the training set, by the architecture of the network, and by the physical complexity of the problem being solved.

In the case considered here there are five input neurons and 16 output neurons. Choosing the appropriate number of hidden neurons is extremely important. To obtain the optimum value of this number, experiments have been performed while keeping other parameters constant. The training started with five hidden neurons; then the number was progressively increased. Figure 6 shows the root mean square error (RMSE) as a function of the number of hidden neurons. It may be observed that a number of 47 hidden neurons gives the best result.

The NN learns about its environment through an iterative process of adjustment. Ideally the network becomes more knowledgeable after each iteration of the learning process. The learning algorithm modifies the weights associated with each processing element to minimize the error between the target and the network actual output. The learning rate is a coefficient that limits the intensity of this modification. High values correspond to fast learning processes, but might cause instability problems. Also the learning rate has been optimized. Figure 7 shows the influence of different learning parameters on the RMSE. The RMSE is given as a function of the epoch size, a number that represents how

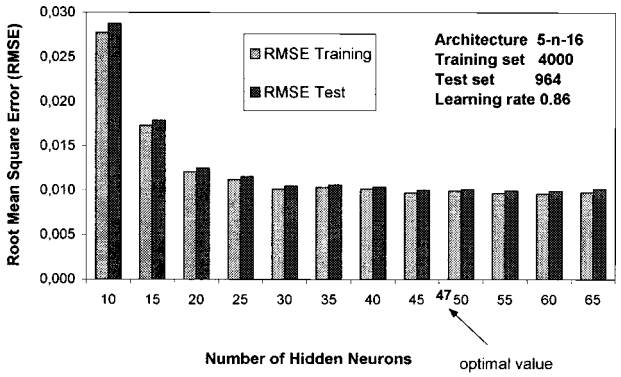


Fig. 6 Optimization of the number of hidden neurons

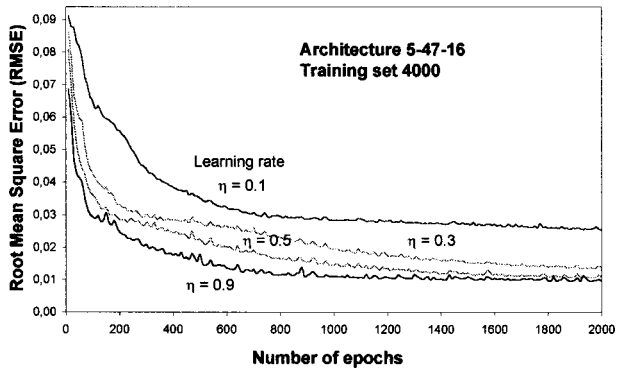


Fig. 7 Optimization of the learning rate value

Table 3 NN optimization configuration

Parameters	Turbo-Gas Section
Number of Input Neurons	5
Number of Hidden Neurons	47
Number of Output Neurons	16
Learning Rate	0.86
RMSE - Training	0.006045
RMSE - Validation	0.006074

many times each data set has been presented to the network. In this application the total number of epochs was 20,000. For the sake of clarity, the range 0–2000 has been magnified. It may be observed that when the learning parameter increases the learning process becomes faster (the curve has a steeper descent). Further, at high numbers of epochs the learning process becomes slower and the curves for learning rates of 0.5, 0.7, and 0.9 are almost the same. A learning rate value of 0.86 has been chosen for all succeeding investigations. The achieved RMSEs for the training and validation sets are 0.006,045 and 0.006,074, respectively. It may be concluded that the learning process is quite fast and the sample size is sufficiently high to give small training and validation (test) errors. The optimized parameters of the trained NN are shown in Table 3.

Validation and Cross Validation. After the optimization of the NN parameters, the results of training were evaluated and compared with the expected values. Then the network was tested to assess its general performance. The trained turbo-gas unit network with optimized parameters has been validated with 964 data sets already used for the training and further verified with 1679 new sets of data. Errors have been calculated as differences between the values obtained by the neural network and the desired ones calculated by the simulator.

The maximum and mean absolute errors during both validation and verification on NN outputs with respect to the desired values are reported in Table 4. It may be observed that the maximum absolute errors are very low, demonstrating good capability in predicting the outputs. An error analysis for the 964 data sets used for the validation is shown in Table 5. Most of the points are very close to the expected values. More than 800 points present errors of almost zero (up to one decimal point) for each NN output variable. Higher errors occur in only a few points.

The trained network was also verified with 1679 new data sets. The absolute error level distribution of NN verification results is shown in Table 6. The absolute errors on fuel mass flow rate, expander exit pressure, waste heat boiler exit pressure, and steam mass flow are zero at all the 1679 points. Some other network output variables show absolute errors lower than one decimal point for about 1500 points. Errors higher than 0.5 are observed at only a few points.

Results of the Neural Network Simulator

Figure 8 shows the variation of gas turbine exhaust temperature, steam temperature, hot water and steam mass flow rates, and fuel mass flow rate as a function of the electric power. Values calculated by the physical simulator, indicated by the lines, are compared with NN values, indicated by circles. The figure is re-

Table 4 Comparison of NN and simulator results

#	Variables	Range of variation	Unit	Maximum Absolute Error		Mean Absolute Error	
				Validation	Verification	Validation	Verification
1	fuel mass flow rate	0.8÷1.7	kg/s	0.0075	0.0088	$0.98*10^{-3}$	$0.967*10^{-3}$
2	air mass flow rate	95÷130	kg/s	0.5123	0.4333	$0.27*10^{-1}$	$0.279*10^{-1}$
3	gas mass flow rate	96÷131	kg/s	0.4797	0.3837	$0.27*10^{-1}$	$0.277*10^{-1}$
4	compressor exit pressure	561÷748	kPa	2.4904	2.6134	0.267	0.291
5	compressor exit temperature	521÷578	K	0.5476	0.6753	$0.67*10^{-1}$	$0.741*10^{-1}$
6	combustion chamber exit pressure	539÷728	kPa	2.1474	2.264	0.252	0.268
7	combustion chamber exit temperature	825÷1110	K	1.6537	1.6552	0.222	0.221
8	blade temperature	825÷1110	K	1.6537	1.6545	0.222	0.221
9	expander exit pressure	90÷104	kPa	0.0703	0.1055	$0.12*10^{-1}$	$0.126*10^{-1}$
10	expander exit temperature	542÷744	K	1.9683	2.3096	0.188	0.195
11	waste heat boiler exit pressure	89÷103	kPa	0.0738	0.1070	$0.11*10^{-1}$	$0.124*10^{-1}$
12	waste heat boiler exit temperature	478÷581	K	2.7685	2.5353	0.166	0.165
13	stack temperature	393÷446	K	0.4636	0.5996	$0.50*10^{-1}$	$0.489*10^{-1}$
14	steam mass flow	0÷8	kg/s	0.126	0.1077	$0.84*10^{-2}$	$0.805*10^{-2}$
15	steam temperature	541÷614	K	2.3085	3.0000	0.321	$0.49*10^{-1}$
16	superheated steam mass flow	1.2÷189	kg/s	2.9333	3.0776	0.250	0.239

Table 5 Results of NN validation

#	Variables	Below 0.1	0.1 ÷ 0.5	0.5 ÷ 1.0	1 ÷ 2	2 ÷ 5	above 5
1	fuel mass flow rate	964	0	0	0	0	0
2	air mass flow rate	942	21	1	0	0	0
3	gas mass flow rate	943	21	0	0	0	0
4	compressor exit pressure	335	504	80	40	5	0
5	compressor exit temperature	777	184	3	0	0	0
6	combustion chamber exit pressure	329	528	68	37	2	0
7	combustion chamber exit temperature	336	540	71	17	0	0
8	blade temperature	336	540	71	17	0	0
9	expander exit pressure	964	0	0	0	0	0
10	expander exit temperature	381	522	55	6	0	0
11	waste heat boiler exit pressure	964	0	0	0	0	0
12	waste heat boiler exit temperature	419	504	32	7	0	0
13	stack temperature	848	116	0	0	0	0
14	steam mass flow	962	2	0	0	0	0
15	steam temperature	232	548	136	45	3	0
16	superheated steam mass flow	323	531	85	22	3	0

Table 6 Results of NN verification

#	Variables	Below 0.1	0.1 ÷ 0.5	0.5 ÷ 1.0	1 ÷ 2	2 ÷ 5	above 5
1	fuel mass flow rate	1679	0	0	0	0	0
2	air mass flow rate	1632	47	0	0	0	0
3	gas mass flow rate	1635	44	0	0	0	0
4	compressor exit pressure	523	883	186	78	9	0
5	compressor exit temperature	1293	376	10	0	0	0
6	combustion chamber exit pressure	543	897	165	70	4	0
7	combustion chamber exit temperature	577	931	145	26	0	0
8	blade temperature	569	940	143	27	0	0
9	expander exit pressure	1679	0	0	0	0	0
10	expander exit temperature	689	865	105	18	2	0
11	waste heat boiler exit pressure	1679	0	0	0	0	0
12	waste heat boiler exit temperature	735	875	48	16	5	0
13	stack temperature	1512	165	2	0	0	0
14	steam mass flow	1679	0	0	0	0	0
15	steam temperature	459	909	229	79	3	0
16	superheated steam mass flow	588	898	153	35	5	0

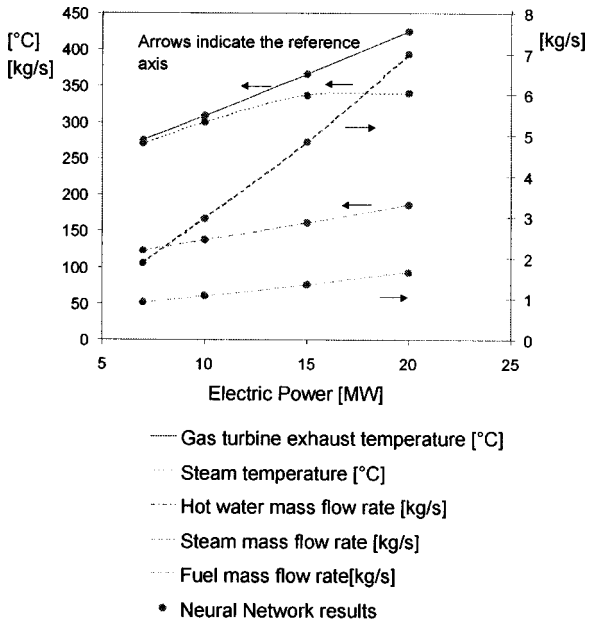


Fig. 8 Waste heat recovery and water heater section

ferred to the ISO conditions, with zero gas bypass. The deterioration and fouling factors have been set at the new status of components.

A very close fit has been achieved between the NN output and results obtained with the physical simulator.

Once the network accuracy has been demonstrated, a very important aspect is the calculation time. A comparison has been performed between the physical simulator and the NN simulator. The ratio of the NN calculation time to the physical simulator calculation time for one operating condition is about one to 1000. The NN computing time was really short.

Conclusions

This work has demonstrated that it is possible to generate maps related to cogeneration plant operating conditions using a NN approach. In order to provide data to train a NN, a numerical prototype of the turbo-gas unit was developed. A method of plant operation planning based on the real simulation of plant components has been proposed. The prototype has shown a good response and prediction capability on components when boundary conditions vary.

A NN simulator was set up and optimized. When the number of hidden neurons increases, the complexity of the NN and the training time increase dramatically. However, the network has shown very encouraging results. The trained NN is capable of predicting

various operating conditions with good accuracy almost instantaneously. From the present investigation the following conclusions can be drawn.

1 The learning is not only affected by the number of input and output variables; the relationships between input and output variables also play an important role.

2 The number of hidden neurons depends not only on the number of input and output neurons, but much more on the relationship between the input and the output. The higher the complexity of the problem, the greater the number of hidden neurons.

3 The NN has predicted almost all the desired output variables with quite good accuracy within the established domain. Therefore, the NN prediction capabilities in gas/steam cogeneration plants seem to be good.

4 A comparison between the NN calculation time and that of the physical simulator has shown a really significant difference in favor of the first. Such a short time allows the utilization of NNs for on-line control purposes and in each case the computing time is an important aspect.

Further investigation should be devoted to the following.

1 Comparison of multilayer NNs with other kinds of existing networks to verify the possibility of improving performance.

2 In order to improve neural network training and to extract more information from the plant different prototypes may be developed. Each prototype may represent a component or a part of the plant.

Acknowledgments

The authors wish to express their gratitude to EU, Università degli Studi Roma Tre, and to MURST, ENEA, and CNR for their support.

References

[1] Parrella, M., Frangopoulos, C. A., Pitt, R. U., Cerri, G., Williams, B. C., Vignolini, M., Psychogios, J., and Salvador, S., 1999, ''Optimum Management System with Environment Monitoring,'' OMSEM Publishable Report, The European Commission Non Nuclear Energy Program, JOULE III.

[2] Cerri, G., Monacchia, S., and Seyedan, B., 1999, ''Optimum Load Allocation in Cogeneration Gas-Steam Combined Plants,'' ASME paper 99-GT-6, International Gas Turbine and Aero Engine Congress and Exhibition, Indianapolis, June 7–10, 1999.

[3] Cerri, G., 1996, ''A Simultaneous Solution Method Based on A Modular Approach for Power Plant Analyses and Optimized Designs and Operations,'' International Gas Turbine and Aeroengine Congress and Exhibition, Birmingham, UK, June 10–13, 1996.

[4] Bohn, D., Dibellius, G. H., Pitt, R. U., Faatz, R., Cerri, G., and Salvini, C., 1992, ''Study on Pressurized Fluidized Bed Combustion Combined Cycles with Gas Turbine Topping Cycle,'' ASME paper 92-GT-343.

[5] Bohn, D., Dibellius, G. H., Pitt, R. U., Faatz, R., Cerri, G., and Salvini, C., 1993, ''Optimizing a Pressurized Fluidized Bed Combustion Combined Cycles with Gas Turbine Topping Cycle,'' ASME paper 93-GT-390.

[6] Cerri, G., 1982, ''Parametric Analysis of Combined Gas-Steam Cycles,'' American Society of Mechanical Engineers, Gas Turbine Division of the ASME, New York, Paper 82-GT-95.

[7] Cerri, G., Sorrenti, A., and Zhang, Q., 1997, ''Optimization of Cleaning Timing and Load Allocation in Steam Generator Management,'' Appl. Therm. Eng., **18**, Nos. 3–4, pp. 111–120.

[8] Hopfield, J. J., and Tank, D. W., 1986, ''Computing With Neural Circuits: A Model Properties Like Those of Two State Neurons,'' Science, **233**, pp. 625–633.

[9] Rumelhart, D. E., McClelland, J. L., and the PDP Research Group, 1987, ''Parallel Distributed Processing: Explorations in the Microstructure of Cognition,'' MIT Press, Cambridge, MA.

[10] Sejnowski, T., and Rosenberg, C. R., 1986, ''NETtalk: A Parallel Network That Learns to Read Aloud,'' Technical Report No. JHU/EECS- 86/01, John Hopkins University, Baltimore, MD.

[11] Boccaletti, C., Cerri, G., Khatri, D. S., and Seyedan, B., 1999, ''An Application Of Neural Network In Combustion Processes Evaluations,'' in Proceedings of International Conference on Enhancement and Promotion of Computational Methods in Engineering and Science (EPMESC VII), Macao, August 2–5, 1999.

[12] Boccaletti, C., Cerri, G., Khatri, D. S., and Seyedan, B., 1999, ''Neural Network Approach in Gas Expansion Evaluations,'' in Fourteenth International Symposium on Airbreathing Engines (ISABE), Florence, Italy, September 5–10, 1999.

A Study of Integrally Augmented State Feedback Control for an Active Magnetic Bearing Supported Rotor System

G. T. Flowers

G. Szász

Department of Mechanical Engineering,
Auburn University,
Auburn, AL 36849

V. S. Trent

M. E. Greene

Department of Electrical Engineering,
Auburn University,
Auburn, AL 36849

State feedback controller designs allow for pole placement in an effective manner, but reduction of static offset is difficult. On the other hand, classical control methodology allows for the increase of system type and the elimination of static offset. An integrally augmented state feedback controller provides the benefits of standard feedback designs while allowing for the elimination of static offsets (through the increase of system type). Static offset is a particular problem with magnetic bearing supported rotor systems, in that gravitational effects, current biasing, and operational loading tend to exacerbate this problem. In order to assess the effectiveness of this technique, an integrally augmented state feedback controller is developed, implemented, and tested for a magnetic bearing supported rotor system. Results for several selected configurations are presented and compared. Some conclusions and recommendations concerning the effectiveness of integrally augmented state feedback controller designs are presented.
[DOI: 10.1115/1.1360686]

Background

The concept of using magnetic media for the suspension of metallic material is not a new one. One of the earliest researchers investigating the magnetic suspension problem was Earnshaw [1] who contemplated the possibility of using permanent magnets for suspension. In his studies, he encountered the inherent instability of the magnetic suspension problem and concluded that suspension of metallic material with permanent magnets was not feasible. This man, although gifted with remarkable insight, was limited by the technology of the times.

Through the 1950s, 1960s, and 1970s advances in semiconductor technology made possible the development of active magnetic (electromagnets) for suspension of metallic media. Researchers at Charles Stark Draper Laboratories at MIT successfully designed, constructed and tested a working magnetic suspension system ([2]). Since the 1970s much research has been focused on using magnetic suspension systems for varied applications ranging from high speed trains to ball valves. One application borne of the advances in silicon based technology and the ever increasing speed of data acquisition systems is the use of active magnetic bearings (AMB) for support of high speed rotors.

High speed turbomachinery applications are often limited by the vibration of the conventional roller bearings used to support the rotating structure (Humphris et al. [3]). Because conventional bearings have fixed characteristics, there is no avenue for active vibration suppression of the rotating structure. AMB, on the other hand, provides a means for vibration suppression of the rotating structure since the characteristics of the bearing may be actively controlled electronically. Initially, efforts in the design of AMB supported rotor systems involved the design of analog circuitry to provide stability for the rotor system. Humphris et al. [3] developed analog controllers using classical design techniques to stabilize an AMB supported rotor. Williams, Keith, and Allaire [4] developed both analog controllers and digital controllers for vibration reduction in a radial AMB system and provided a comparison of the results. Eventually, the design techniques employed to develop the controllers for AMB supported rotor systems advanced from the classical techniques to the use of state-space design techniques. Burrows et al. [5] presented a pole-placement controller for a flexible rotor system using state-space analysis techniques. Hung [6] developed a nonlinear controller using feedback linearization and compared the results to a PD controller design. More recently, Shafai et al. [7] present a study of the synchronous vibration problem along with an overview of LQG/LQT, H_∞, and quantitative feedback theory.

The development of state-space based control methodologies provides an effective avenue for pole placement of the compensated system. The judicious selection of closed-loop poles, in turn, provides the engineer with the ability to develop desirable transient response characteristics for the compensated system. However, the reduction or elimination of static offset is difficult. Classical control methodologies, on the other hand, provide an effective avenue for the reduction or elimination of static offsets through increasing the system type (i.e., PI and PID controllers). Researchers have combined the two methodologies through the design of state–feedback control systems employing integral augmentation. Czarkowski and Kazimierczuk [8] used integral control with state feedback to control a pulse-width modulated push-pull dc converter. Saif [9] employed a state feedback approach with integral action to control a multivariable servomechanism. Additionally, the idea of using integral compensation in conjunction with sliding mode or variable structure control has been investigated by Chung et al. [10] and by Ho and Sen [11].

This paper presents the modeling, design, implementation and testing of an integrally augmented state-feedback controller for use in an active magnetic bearing supported rotor. First, descriptions of the experimental model and the simulation model are presented. Next, the controller design methodology is described and the controller design is implemented and tested. Results are then presented and discussed for some selected parametric configurations. Finally, some conclusions and recommendations con-

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Orlando, FL, June 2–5, 1997; ASME Paper 97-GT-231. Manuscript received by IGTI December 1996; final revision received by the ASME Headquarters Mar. 1997. Associate Editor: H. A. Kidd.

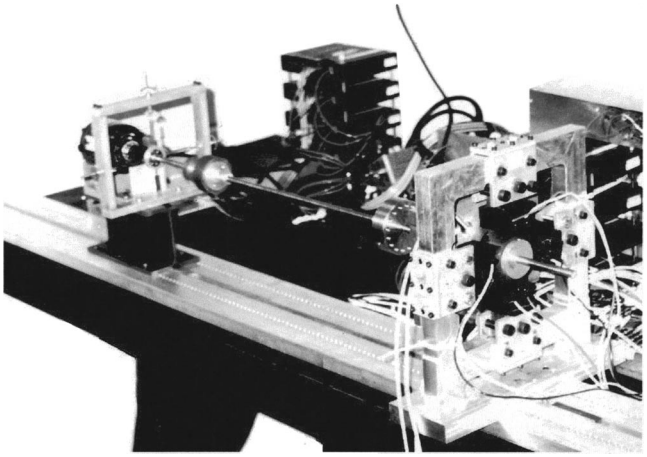


Fig. 1 Photograph of experimental apparatus

cerning the effectiveness of an integrally augmented state feedback control system for use in an active magnetic bearing supported rotor system are presented.

Experimental Model

Experimental tests were performed in order to validate the behavior predicted by the simulation model and to gain some insight into the dynamical responses that are to be expected. A photograph of the experimental test rig is shown in Fig. 1. It has two basic components: a flexible shaft and an auxiliary clearance bearing rig. The shaft is made of steel and is 9.8 mm in diameter and 0.91 m in length. It is supported at the left end by a ball bearing suspended in a frame by four springs and at 76.2 mm from the right end by a magnetic bearing. The stiffness of the left support, for both horizontal and vertical directions, is 17,000 N/m. A rigid disk with threaded boreholes for placing imbalance screws is positioned at the midpoint of the bearing span.

The rotor is driven by an adjustable speed motor with feedback speed controller. Shaft vibration is measured using four eddy current proximity sensors fixed to measure displacements. The sensors are fixed at 90 degrees from each other and 45 degrees from the vertical axis. The signals from these sensors are used to calculate the horizontal and vertical displacements of the rotor.

Simulation Model

Figure 2(*a*) shows the simulation model used for this investigation. It consists of a flexible rotor supported by a magnetic bearing at one end and springs at the other end. A block diagram for the system is shown in Fig. 2(*b*). The rotor is modeled using the bending mode shapes and natural frequencies (obtained through finite element analyses) for a shaft supported by a spring constraint (spring constant is 17,000 N/m) at the left end and free at the right end. The finite element code uses 73 stations and the first four modes (two rigid body and two flexible modes) are included in the simulation model. The rotor equations of motion can then be written as:

$$\ddot{\mathbf{q}}_y + \Omega\Gamma\dot{\mathbf{q}}_x + \omega_n^2\mathbf{q}_y = \phi_T F_{y,amb} + \phi^T F_{y,lin} + \phi^T F_{y,imb}, \quad (1a)$$

$$\ddot{\mathbf{q}}_x - \Omega\Gamma\dot{\mathbf{q}}_y + \omega_n^2\mathbf{q}_x = \Phi_T F_{x,amb} + \Phi^T F_{x,lin} + \Phi^T F_{x,imb}, \quad (1b)$$

where

$$\mathbf{F}_{y,lin} = \Phi_{N_{b1}} C_y \dot{\mathbf{q}}_y$$

$$\mathbf{F}_{x,lin} = \Phi_{N_{b1}} C_x \dot{\mathbf{q}}_x$$

$$\mathbf{F}_{y,imb} = \psi\Omega^2 \sin\Omega t,$$

$$\mathbf{F}_{x,imb} = \psi\Omega^2 \cos\Omega t,$$

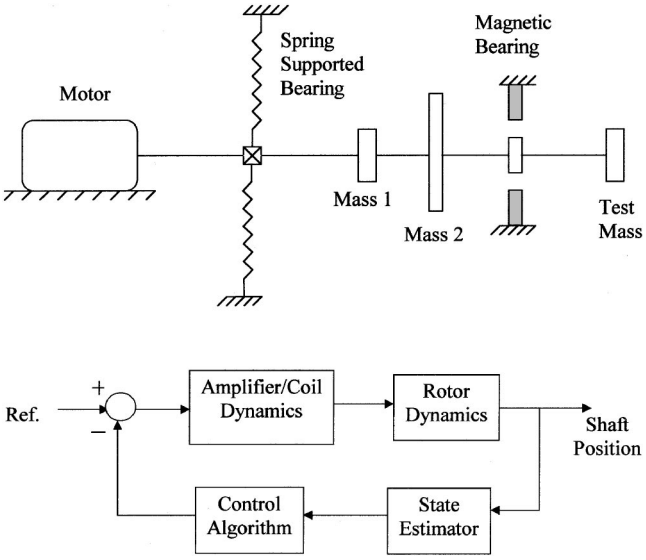


Fig. 2 (*a*) Schematic diagram of simulation model, (*b*) block diagram of model system

$$\mathbf{F}_{y,amb} = \mu_o a N_T^2 \left(\frac{I_1^2}{(d-y_b)^2} - \frac{I_2^2}{(d+y_b)^2} \right),$$

$$\mathbf{F}_{x,amb} = \mu_o a N_T^2 \left(\frac{I_3^2}{(d-x_b)^2} - \frac{I_4^2}{(d+x_b)^2} \right),$$

$$I_1 = I_1^B + i_1,$$

$$I_2 = I_2^B + i_2,$$

$$I_3 = I_3^B + i_3,$$

$$I_4 = I_4^B + i_4,$$

$$I_1^B = K_a V_1^B,$$

$$I_2^B = K_a V_2^B,$$

$$I_3^B = K_a V_3^B,$$

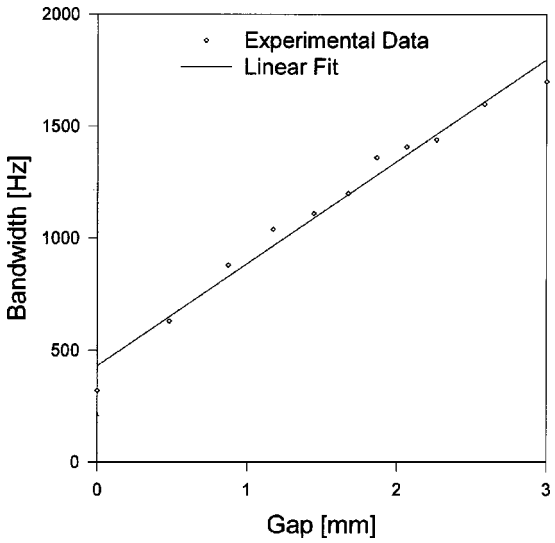


Fig. 3 Amplifier/coil bandwidth as a function of magnetic bearing gap

$$I_4^B = K_a V_4^B,$$

The amplifiers used with the magnetic bearing are pulse width modulation (PWM) amplifiers of moderate bandwidth. Experimental testing has shown that the bandwidth is (approximately) a linear function of the relative gap between the soft iron core of the magnetic bearing and the stator coils. Some experimental data and the linear curve fit are shown in Fig. 3. Based upon this testing, the governing differential equations for the coil/amplifier dynamics are:

$$\dot{I}_1 + -(h_1 + h_2(d - x_b))I_1 = K_A(h_1 + h_2(d - y_b))V_1 \quad (1c)$$

$$\dot{I}_2 + (h_1 + h_2(d + y_b))I_2 = K_A(h_1 + h_2(d + y_b))V_2 \quad (1d)$$

$$\dot{I}_3 + (h_1 + h_2(d - x_b))I_3 = K_A(h_1 + h_2(d - x_b))V_3 \quad (1e)$$

$$\dot{I}_4 + (h_1 + h_2(d + x_b))I_4 = K_A(h_1 + h_2(d + x_b))V_4 \quad (1f)$$

The physical displacements of the rotor at the magnetic bearing location can be obtained using the following coordinate transformation:

$$x_b = \sum_{i=1}^{N} \Phi_{N_{b,i}} Q_{x,i},$$

$$y_b = \sum_{i=1}^{N} \Phi_{N_{b,i}} q_{y,i},$$

Controller Design

The governing equations for the model system, as shown above in (1), are highly nonlinear. In order to design a controller/estimator for the system, a reduced order linearized model is employed. For the rotor, only the first mode (rigid body rotation about the left spring support) is considered and the governing equations are linearized about the bearing center position and the respective bias currents. In addition, the individual coil currents are not observable. However, inspection of the governing equations shows that they can be manipulated so that the individual coil currents are replaced with the new variables $I_1^B i_1 - I_2^B i_2$ and $I_3^B i_3 - I_4^B i_4$. The resulting system of equations is completely observable.

$$\dot{\overline{x}}_D = A_D \overline{x} D + B_D \overline{u}_D \quad (2a)$$

$$\overline{y}_D = C_D \overline{x}_D \quad (2b)$$

$$A_D = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ K_y & 0 & -G_y & \omega\Gamma_1 & K_i & 0 & 0 & 0 \\ 0 & K_x & -\omega\Gamma_1 & -G_x & 0 & K_i & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{K_A}{\tau} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\frac{K_A}{\tau} & 0 & 0 \\ \phi_{1,N_b} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \phi_{1,N_b} & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$B_D = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ \frac{K_A}{\tau}(I_1^B + I_2^B) & 0 \\ 0 & \frac{K_A}{\tau}(I_3^B + I_4^B) \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$C_D = \begin{pmatrix} \phi_{1,N_b} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \phi_{1,N_b} & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

where:

$$K_y = \frac{2a\mu_o N_T^3 (I_1^{B^3} + I_2^{B^3})\phi_{1,N_b}}{d^3}$$

$$K_x = \frac{2a\mu_o N_T^3 (I_3^{B^3} + I_4^{B^3})\phi_{1,N_b}}{d^3}$$

$$K_i = \frac{2a\mu_o (nt^2)\phi_{1,N_b}}{d^2}$$

$$G_y = \phi_{1,n_b}^2 c_y$$

$$G_x = \phi_{1,n_b}^2 c_x$$

$$\tau = \frac{1}{2\pi(h_1 + h_2 d)}$$

The design states and the model states are related as:

Table 1 Simulation model parameters

Parameter	Value	Units
a	3.22×10^{-4}	m^2
c_x, c_y	75.0	N s/m
h_1	2.70×10^{3}	1/s
h_2	2.87×10^{6}	1/s m
K_a	0.42	amps/V
N_T	160	
V_1^B	4.0	volts
V_2^B	2.0	volts
V_3^B	3.5	volts
V_4^B	3.5	volts
d	1.5×10^{-3}	m
ψ_{48}	2.5×10^{-5}	kg m
μ_o	1.26×10^{-6}	H/m

$$\overline{x}_D = \begin{pmatrix} q_1^x \\ q_1^y \\ \dot{q}_1^x \\ \dot{q}_1^y \\ I_1^B i_1 - I_2^B i_2 \\ I_3^B i_3 - I_4^B i_4 \\ x_1^b \\ y_1^b \end{pmatrix}$$

The state estimator is designed using the first six states. The last two (integral states) are calculated directly from the measured position data. Using the estimated states, the control law can be formulated as:

$$\overline{u} = -K_c \overline{x}_e \tag{3}$$

Pole placement is used to select the feedback gains K_c for this study and the base values that were used are:

$$\begin{pmatrix} 300+j\sigma_P & 300-j\sigma_P \\ 300+j\sigma_P & 300-j\sigma_P \\ \sigma_I & \sigma_I \\ -10{,}000 & -10{,}000 \end{pmatrix}$$

Please note that identical values are used for the poles in both the horizontal and vertical directions. Values for σ_P and σ_I are varied in the studies that are presented. Specific values are shown on the respective figures. Experience with the model has shown that σ_I is the primary pole for the selection of the integral feedback gains, σ_P is the primary driver for the selection of the proportional feedback gains, and the last two poles ($-10{,}000$ values are shown) are the primary drivers in the selection of the current feedback gains (see Table 1).

Discussion

The controller/estimator design described above was implemented on the experimental facility and tested. The primary objective of the tests were to evaluate the robustness of the controller design and the effectiveness of integral augmentation in reducing steady-state errors. A major source of such effects can result from modeling errors and unexpected sideloading. In order to simulate a (large) modeling error and provide a substantial (gravity) sideloading, a mass of 0.82 kg was added at the right end of the rotor shaft. The test procedure consists of adding the mass, supporting the extra weight with a strap around the shaft, activating the magnetic bearing control system, allowing transient effects to die down, and removing the strap support. The dynamic behavior of the shaft before and after the release of the support strap were observed and recorded. Test results for selected cases are shown below.

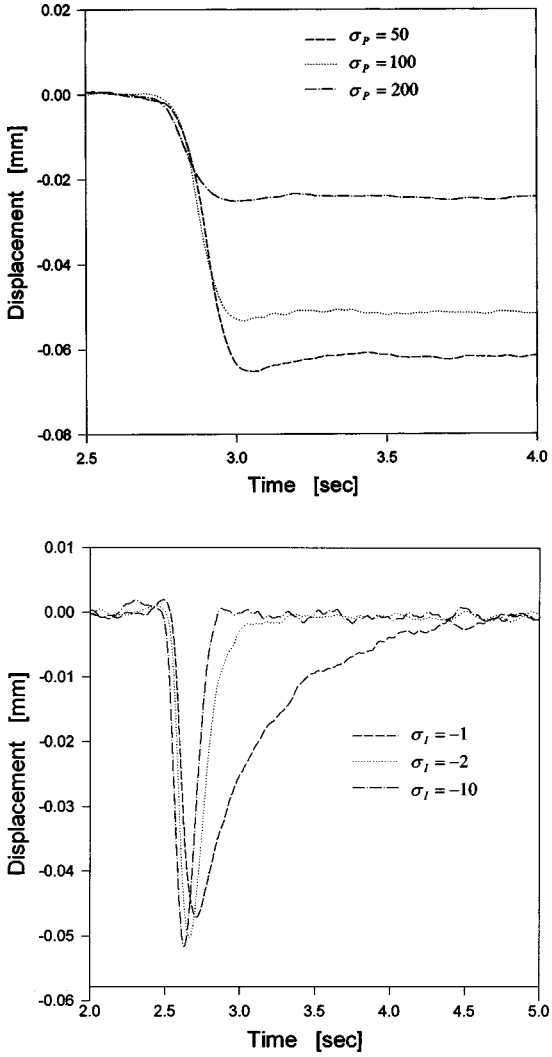


Fig. 4 (*a*) Experimental rotor response for the nonspinning system without integral augmentation ($\sigma_I = 0$), (*b*) experimental rotor response for the nonspinning system with integral augmentation ($\sigma_P = 100$)

Figures 4*a* and 4*b* show the experimentally observed vibrational behavior for the nonspinning rotor. Similar behavior is observed for the corresponding simulation responses. For the system with no integral augmentation, the rotor drops and oscillates for a few cycles before settling at a new equilibrium position that is substantially below the original operating position. As σ_P is increased, the effective stiffness of the bearing increases and the steady–state position offset decreases accordingly. For the system with integral augmentation, the rotor drops, but recovers in an exponential fashion. There is some oscillation as the equilibrium position is approached, but the final result is no steady-state offset. As σ_I is increased, the integral action of the controller increases, resulting in a faster return to the nominal operating position.

For the spinning rotor, similar results are observed. Figures 5*a* and 5*b* show the experimentally observed vibrational behavior for the rotor spinning at 90 Hz. Again, similar behavior is observed for the simulated responses. The controller was designed to allow for the inclusion of gyroscopic effects, but they are negligible for this system. For the system with no integral augmentation, the rotor settles rapidly into a new (and lower) equilibrium position. For the system with integral augmentation, the rotor again recovers in an exponential fashion with no discernable steady-state offset.

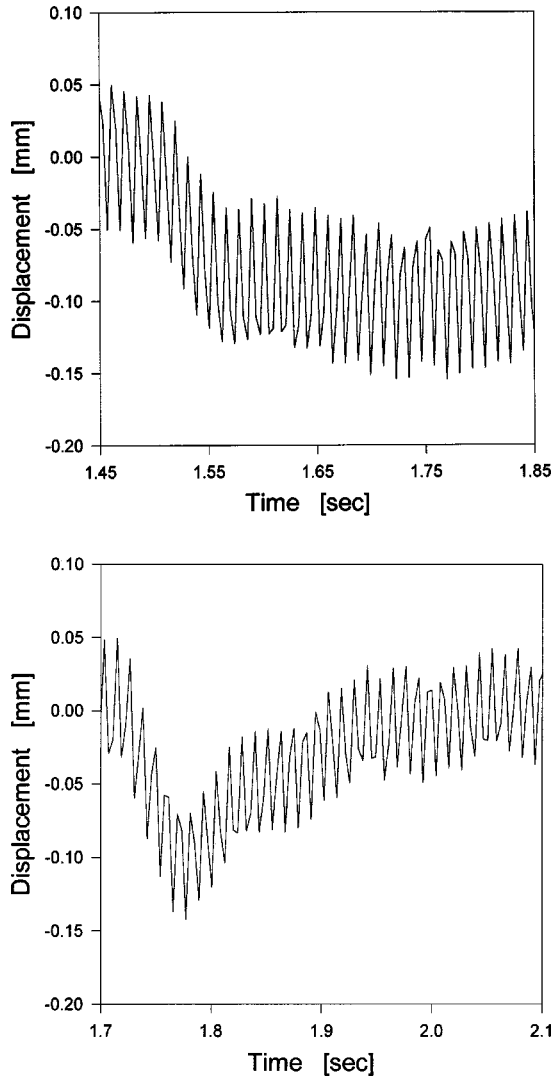


Fig. 5 (*a*) Experimental rotor response for the spinning system without integral augmentation ($\sigma_I=0$, $\sigma_P=100$), (*b*) experimental rotor response for the spinning system with integral augmentation ($\sigma_I=-10$, $\sigma_P=100$)

Summary and Conclusions

An integrally augmented controller for a flexible rotor system was designed, implemented, and tested. The objective was to illustrate the methodology for the development of such a controller and to test its effectiveness. A model for a flexible shaft rotor system with a magnetic bearing supporting one end was developed and used to design a controller/estimator for the system. A nonlinear model for the bearing amplifier/coil dynamics was identified and used in the system model. Simulation studies and experimental tests were conducted. The system responds very well to a linear controller, even though it is, in fact, quite nonlinear. It was found that an integrally augmented state feedback controller/estimator was very successful in providing good transient dynamic control and the virtual elimination of steady-state offset.

Acknowledgments

This work was supported by NASA under Grant No. NAG3-1507 and by NSF under Grant No. NSF-CMS-9503488. The government has certain rights in this material. Appreciation is expressed to A. F. Kascak and D. P. Garg for their advice and assistance in this research effort.

Nomenclature

a = pole face area
A_D = estimator linear state matrix
B_D = estimator input matrix
C_D = estimator output matrix
c_y = rotor damping in Y direction, N sec/m
c_x = rotor damping in X direction, N sec/m
d = radial clearance in magnetic bearing, m
h_1, h_2 = amplifier bandwidth linear constants
K_y = elastic stiffness in Y direction, N/m
K_x = elastic stiffness in X direction, N/m
I_a = rotor polar mass inertia matrix, kg m^2
I_j^B = bias current for the ''jth'' coil, amps
i_j = control current for the ''jth'' coil, amps
N = total number of modes considered
N_b = node number at rightmost bearing
N_T = number of coil turns per leg
$\mathbf{q}_x$ = rotor modal coordinate vector in X direction
$\mathbf{q}_y$ = rotor modal coordinate vector in Y direction
t = time, s
$\overline{u}$ = state feedback control law
V_j^B = bias voltage for the ''jth'' coil, volts
x_b = rotor displacement at magnetic bearing in X direction, m
y_b = rotor displacement at magnetic bearing in Y direction, m
$\overline{x}$ = state vector
x_I^b = integral of rotor displacement at magnetic bearing in X direction, m
y_I^b = integral of rotor displacement at magnetic bearing in Y direction, m
$\overline{x_e}$ = estimated state vector
Ψ = rotor spring-supported/free modal rotation matrix
Φ = rotor spring-support/free modal displacement matrix
Γ = $\Psi^T I_a \Psi$
Ω = rotor operating speed, rad/s
μ_o = permeability of free space, H/m
ω_n = matrix of rotor free–free natural frequencies, rad/s
ψ = imbalance vector, kg m

Subscripts

$(\dot{})$ = differentiation with respect to time
$\underline{[\,]}$ = matrix quantity
$\underline{(\,)}$ = indicates a vector quantity

References

[1] Earnshaw, S., 1842, ''On the Nature of the Molecular Forces Which Regulate the Constitution of the Luminiferous Ether,'' Trans. Cambridge Philos. Soc., **7**, pp. 97–112.
[2] Frazier, R. H., 1974, *Mechanical and Electrical Suspensions*, The MIT Press, Cambridge, MA.
[3] Humphris, R. R., Kelm, R. D., Lewis, D. W., and Allaire, P. E., 1986, ''Effect of Control Algorithms on Magnetic Journal Bearing Properties,'' J. Eng. Gas Turbines Power, **108**, pp. 624–632.
[4] Williams, R. D., Keith, F. J., and Allaire, P. E., 1991, ''A Comparison of Analog and Digital Controls for Rotor Dynamic Vibration Reduction Through Active Magnetic Bearings,'' J. Eng. Gas Turbines Power, **113**, pp. 535–543.
[5] Burrows, C. R., Keogh, P. S., and Tasaltin, R., ''Closed-loop Vibration Control of Flexible Rotors—An Experimental Study,'' *Proceedings of the Institute of Mechanical Engineers, Part C., Journal of Mechanical Engineering Science*, **207**, pp. 1–17, 1993.
[6] Hung, J. Y., ''Nonlinear Control of a Magnetic Leviation System,'' 1991 International Conference on Industrial Electronics, Control and Instrumentation, Kobe, Japan.
[7] Shafai, B., Beale, S., LaRocca, P., and Cusson, E., 1994, ''Magnetic Bearing Control Systems and Adaptive Force Balancing,'' IEEE Control Syst., Apr., pp. 4–13.

[8] Czarkowski, D., and Kazimierczuk, M. K., 1994, ''Application of State Feedback with Integral Control to Pulse-Width Modulated Push-Pull DC-DC Convertor,'' IEE Proc.-D: Control Theory Appl., **141**, No. 2, Mar., pp. 99–103.
[9] Saif, M., 1993, ''Reduced Order Proportional Integral Observer With Application,'' J. Guid. Control, Dyn., **16**, No. 5, pp. 985–988.
[10] Chung, S. K., Lee, J. H., Ko, J. S., and Youn, M. J., 1995, ''Robust Speed Control of Brushless Direct-Drive Motor Using Integral Variable Structure Control,'' IEE Proc. Electr. Power Appl., **142**, No. 6.
[11] Ho, Edward Y. Y., and Sen, P. C., 1991, ''Control Dynamics of Speed Drive Systems Using Sliding Mode Controllers with Integral Compensation,'' IEEE Trans. Ind. Appl., **27**, No. 5, Sept.–Oct., pp. 883–892.

Improving the Stability of Labyrinth Gas Seals

K. Kwanka
Chair of Thermal Power Systems,
Technische Universität München,
Munich, Germany

The flow through labyrinth seals of turbomachinery generates forces which can cause self-excited vibrations of the rotor above the stability limit. The stability limit is reached at a specific rotating speed or power. The continuous growth of power density and rotating speed necessitates an exact prediction of the stability limit of turbomachinery. Usually the seal forces are described with dynamic coefficients. A new, easy-to-handle identification procedure uses the stability behavior of a flexible rotor to determine the dynamic coefficients. Systematic measurements with a great number of labyrinth seal geometries lead to reasonable results and demonstrate the accuracy and sensitivity of the procedure. A comparison of the various methods used to minimize the excitation indicates which seal is more stable and will thus improve the dynamic behavior of the rotor.
[DOI: 10.1115/1.1359772]

Introduction

High performance turbomachinery shows a continuous trend towards higher power density and better efficiency. One of the best methods to increase efficiency is by reducing the leakage losses. This can be achieved by decreasing the clearance between the rotating and the nonrotating parts, possibly combined with new geometries of the cavity and/or a somewhat roughened stator surface.

On the other hand, fluid-induced forces in labyrinth seals which are possibly increased by the hydrodynamic bearing excitation can cause unstable vibrations of the rotor. Therefore an accurate and reliable prediction of the working mechanisms and the magnitude of the exciting forces is necessary. This objective is mainly achieved by influencing the flow in the seal, especially the flow in circumferential direction. The use of active elements as magnetic bearings, that would avoid such kind of stability problems, are considered only sporadically.

The fluid-rotor interaction is usually described in the following linear reaction-force/motion model, which uses direct and coupled stiffness and damping coefficients ([1]). In gas seals the added-mass terms are negligible.

$$\vec{F} = -\begin{pmatrix} K & k \\ -k & K \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} C & c \\ -c & C \end{pmatrix}\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} \tag{1}$$

The above equation assumes small motions around the centered position, but even for small deviations out of this position the coefficients in the matrices remain skew-symmetric ([2]). The linearity of the equation can be assumed up to an eccentricity ratio of 0.5. The destabilizing force caused by the cross-coupled stiffness is counteracted by the direct damping. In case of instability, the vibrations due to unbalance are overlaid with the vibration of a large amplitude at the lowest natural frequency.

The labyrinth seal forces result from a specific circumferential pressure distribution. The experimental determination of the forces can be performed in two ways: either by a direct measurement of the forces acting on the bearings or the casing or indirectly considering the pressure distribution. When damping coefficients are required in addition to the stiffness coefficients, then a motion of the rotating rotor relative to the casing is necessary.

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Orlando, FL, June 2–5, 1997; ASME Paper 97-GT-232. Manuscript received by IGTI Dec. 1996; final revision received by the ASME Headquarters Mar. 1997. Associate Editor: H. A. Kidd.

The forces developed in gas labyrinth seals are smaller than in liquid seals by at least one order in magnitude; therefore, only few experimental results are available. Childs [1] gives a survey of the experimental approaches and the investigated seal geometries. All authors use very stiff and heavy rotors, which are forced to perform a motion by means of more or less complicated mechanisms with hydraulic ([3–5]) or electromagnetic ([6]) shakers. Other authors employ shaft-in-shaft arrangements ([7]) to decouple the rotating from the whirling motion of the rotor.

In still other new concepts the stiff rotor is carried in active magnetic bearings ([8,9]). Meanwhile Wagner and Steff [2] have succeeded in carrying out first measurements with gas seals by using the latter cited test rig, which is designed for high pressures and rotating speeds.

Identification Procedure and Test Rig

A different approach employs a flexible rotor and uses the change in the dynamical behavior caused by the flow through a labyrinth seal for identification purposes ([10]). The flow-induced forces change the stiffness of the rotor and implicitly the frequency of the self-excited vibrations. On the other hand, the excitation and damping of the seal exert an influence on the stability of the system. Assuming a circular whirling orbit, the unknown dynamic coefficients of Eq. (1) can be expressed by the measurable changes Δr and Δq from the initial values in the dynamic system.

$$\Delta r = -K - c\Omega \tag{2}$$

$$\Delta q = k - C\Omega \tag{3}$$

The measurement of the changes is performed at the stability limit of the rotordynamic system. The excitation or damping of the rotor is obtained by means of a magnetic bearing that is placed on the rotor in addition to the test-seal. With the magnetic bearing, normal and cross-forces can be applied to the rotor. Note, that the magnetic bearing is not used as a bearing in the traditional sense and therefore needs no highly-sophisticated and expensive control system. The actual use of a magnetic bearing was inspired by the work of Ulbrich [11].

The value of Δr is given by the magnetic stiffness necessary to compensate the flow-related whirl frequency variation. The change in magnetic excitation, which is needed to destabilize or stabilize the rotor with and without flow through the test-seal q and q_0, represents Δq (see Fig. 1).

By repeating the measurement of Δr and Δq for two whirling frequencies, the stiffness and damping coefficients can be calculated. It turns out that the accuracy of the damping coefficients is

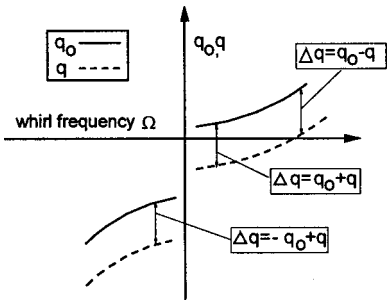


Fig. 1 Change in stability of the forward and the backward mode caused by the seal

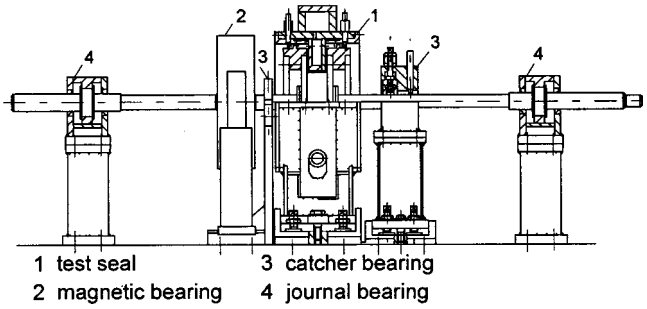


Fig. 2 Test rig for the identification of dynamic coefficients

improved considerably when measurements are carried out at backward whirling frequencies, too. The dynamic coefficients in this paper are identified by a least square fit using the changes in dynamic behavior at minimum three forward and three backward whirling frequencies.

The test apparatus employed in numerous identification runs is displayed in Fig. 2. The diameter of the rotor carried in hydrodynamic bearings is 23 mm. The spacing between the bearing and, accordingly, the whirling frequency varies from 770 to 890 mm or 26 to 35 Hz, respectively. The test-seals are located by pairs symmetrically to the inflow region in the midspan position of the shaft. The prerotation of the flow entering into the seals is generated with a guide vane ring. High amplitudes of the unstable rotor are avoided with catcher bearings.

As the magnetic bearing is not placed at the same axial position as the test-seal, a conversion procedure is used ([12]). The simple and experimentally validated procedures use the amplitudes of the bending modes. In addition, also the influence on the bending mode shape is considered which is caused by the compensation of the whirl frequency shift.

Seal Geometry and Accuracy of Measurement

Both the test-seal internals and the guide vane ring can be easily removed and changed to other labyrinth geometries or guide vane angles. As a reference for all following measurements, a see-through labyrinth is used in teeth-on-rotor configuration (TOR29, tooth width 0.8 mm) with 9 cavities. The pitch is 4 mm and the cavity height is 3.75 mm. The clearance, located at a diameter of 180 mm, is held constant at 0.5 mm for all geometries. In Table 1 the nomenclature of the seal geometry is summarized.

The entry swirl and the circumferential flow in the cavities have a great influence on the magnitude of the generated forces. Therefore, a number of investigations were carried out to improve the rotordynamic response by disturbing the circumferential flow. Leie and Thomas [13] and Benckert and Wachter [14] used swirl brakes mounted in the cavities or in front of the seal. An almost

Table 1 Nomenclature of the used seal geometries

Abreviation	Specification
TOR29	teeth on rotor, smooth stator, 9 cavities (pitch 4 mm, height 3.75 mm)
TOR294DB	teeth on rotor, smooth stator, 4 swirl brakes
TOR298DB	teeth on rotor, smooth stator, 8 swirl brakes
TOR293HW	teeth on rotor, honeycomb stator, cell depth 3 mm
TOR2938H	teeth on rotor, honeycomb stator, cell depth 3 mm, 8 swirl brakes

total reduction of the lateral force is reported by the latter authors, when 8 swirl webs are placed at the entrance of the labyrinth.

In the present paper, an arrangement with 4 (4DB) and with 8 (8DB) swirl brakes in front of the reference seal (TOR29) is investigated (see Fig. 3). The best known use of swirl brakes was reported in the high-pressure-fuel turbopump of space shuttle ([15]). In this application a honeycomb-stator is inserted in the seal beside the swirl brake.

In general, it is expected that honeycomb seals show favorable dynamic characteristics and, last but not least, smaller leakage rates. First investigations of Hawkins et al. [4] indicate that a 16-cavity TOR seal does not improve the stability significantly. Unfortunately the number of cavities, the honeycomb cell width and the pressure levels differ and therefore a direct comparison with the actual results is not possible. When the reference seal (TOR29) is used with a honeycomb-stator, the leakage rate is reduced to 60 percent of the initial value. As to isolate the effects, the seal is first used without swirl brakes. The main dimensions of the honeycombs are mentioned in Fig. 4.

As mentioned before, the seal geometry and the guide vane ring can easily be removed and changed. The preswirl, which is of outstanding importance for the generation of the exciting forces, is varied by using guide vanes with different nozzle angles. The magnitude of the entry swirl is estimated out of the metered mass-flow through the seals and the nozzle area and angle. This is maybe only an approximation of the real entry swirl, but the inflow situation in the real axial turbomachinery is inhomogeneous, too, and therefore probably very similar. On the other hand the geometry of the seal entrance is held unchanged for all test runs and, accordingly, the comparison of different seal concepts remains absolutely reliable.

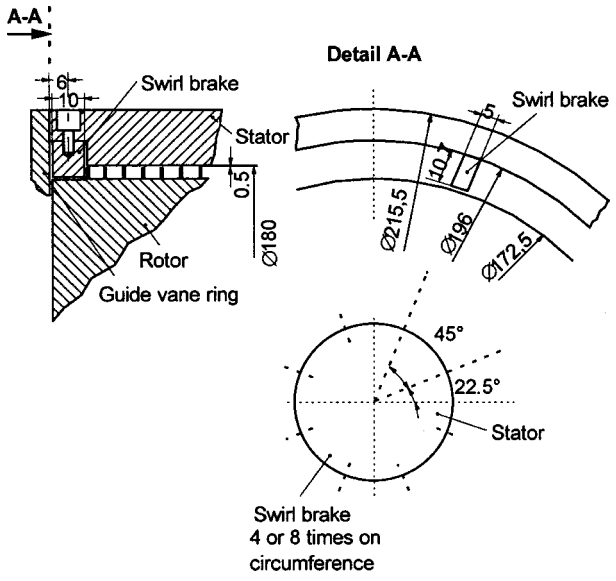


Fig. 3 Arrangement of swirl brakes at the entrance of the seal

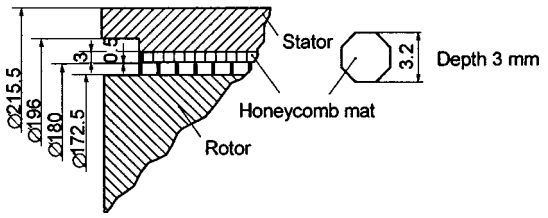


Fig. 4 See-through labyrinth with honeycomb-stator

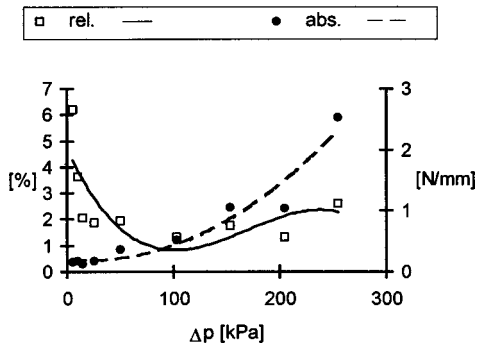


Fig. 5 Relative and absolute deviation between measurement and least square fit

Table 2 Reproducibility of identified dynamic coefficients

Measurement	k [N/mm]	C [Ns/mm]	K [N/mm]	c [Ns/mm]
1	93.3	.0592	32.8	.0151
2	91.6	.0574	33.9	.0180

The accuracy of the procedure is determined by the clear identification of the stability limit. With very few exceptions, the rotordynamic stability limit can be fixed within ±1–2 N/mm. The deviations of the measured points from the best fitted line are usually smaller than 1.5 percent ([16]). Only with low pressure differences and therefore small seal forces are the relative deviations higher. When the forces are stronger and the vibrational behavior rougher, then the absolute deviations are high. The typical course of the absolute and relative deviation in dependence of the pressure difference is illustrated in Fig. 5. Of great benefit for the accuracy and sensitivity of the procedure is the arrangement with two seals in parallel that helps simultaneously double and average the forces.

The position of the journal in the hydrodynamic bearing remains the same with and without flow through the test seal; in consequence, the bearings are expected to have any influence on the results. Additionally, this is confirmed by the good reproducibility of the measurements. All four dynamic coefficients of a see-through labyrinth with 4 cavities remain nearly the same after a total disassembly and reassembly of the test rig (Table 2), despite an assumptive changed alignment of the bearings and therefore a somewhat altered behavior of the bearing.

Experimental Results

Previous measurements have demonstrated the expected strong dependency of the cross-coupled stiffness with the entry swirl and the pressure difference ([16]). However, the influence of the rotating speed is small in the investigated range up to a maximum rotor surface velocity of 47 m/s. Next, the effectiveness of measures to minimize the exciting forces is investigated.

The dynamic coefficients of a see-through seal with 9 cavities used as reference are displayed in Fig. 6. The cross-coupled stiffness shows the usual almost linear dependency on the preswirl.

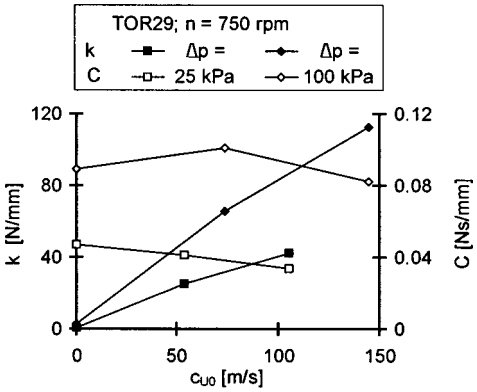


Fig. 6 Cross-coupled stiffness and direct damping of a see-through labyrinth seal with 9 cavities

The behavior of the direct damping is distinctive different. Even without entry swirl, the damping values differ from zero and do not depend on the preswirl. With an increased pressure difference, the leakage rate and the preswirl are higher, too.

In a next step, swirl brakes are placed at the entrance of the seal. The swirl brakes are equally spaced on the circumference as illustrated in Fig. 3. Four swirl brakes suffice to reduce the cross-coupled stiffness by more than 50 percent (see Fig. 7).

Eight brakes on the circumference allow an even further reduction. It is remarkable to see that the direct damping is increased simultaneously. In case of zero preswirl and small drag due to rotation, there is no difference between the measurement with and without brakes. The swirl brakes have no influence on the leakage.

Now the swirl brakes are removed and the stator of the seal is replaced by a honeycomb stator. Again the cross-coupled stiffness is reduced, although not as strongly as with swirl brakes. At the same time the direct damping is decreased, too (see Fig. 8).

Because of the smaller leakage with the honeycomb stator the dynamic coefficients of the reference seal are converted to the

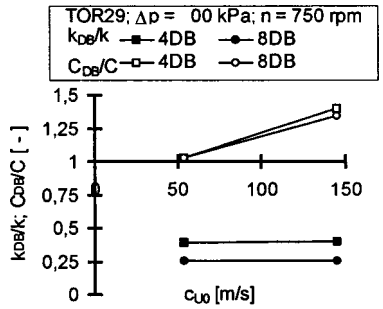


Fig. 7 The effect of swirl brakes on the dynamic coefficients (*i*DB-*i* number of swirl brakes on circumference)

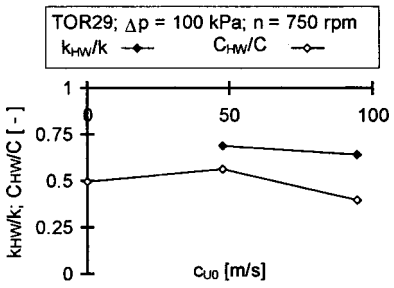


Fig. 8 The effect of a honeycomb stator on the dynamic coefficients (HW: honeycomb)

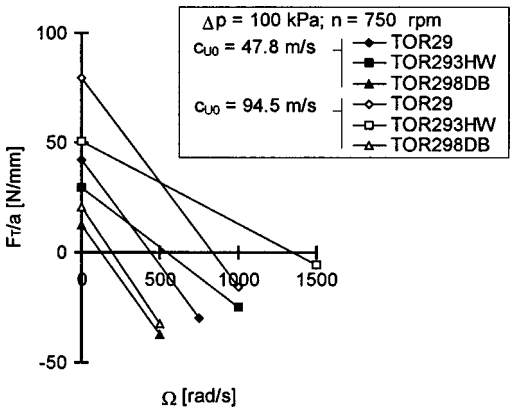


Fig. 9 Exciting force in dependence of whirling frequency and preswirl

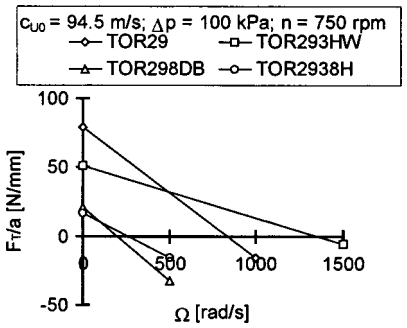


Fig. 10 Minimizing the exciting forces with swirl brakes

same preswirl values. In both cases the cross-coupled stiffness with negligible swirl is very small. Therefore, a comparison does not make sense.

Now the question is, which labyrinth seal concept is more suited from the rotordynamical point of view. The comparison of the lateral force F_T divided by the whirling amplitude a of the circular modal vibration indicates the magnitude of the excitation at a specific whirling frequency Ω. The excitation can be calculated by inserting the actual dynamic coefficients in Eq. (2). A negative excitation indicates that the seal contributes to the stability of the system.

The comparison of the actual seal concepts for two preswirl velocities is displayed in Fig. 9. The cross-coupled stiffness is given by the intersection with the ordinate and increases along with the swirl. Due to the direct damping the excitation is reduced with increasing whirling frequency. The see-through seal employing swirl brakes shows by far the lowest excitation. The honeycomb seal is favorable only with lower whirl frequencies, whereas at higher whirl frequencies the see-through labyrinth without swirl brakes seems to be advantageous. In Fig. 9 again a conversion of the dynamic coefficients to a common basis in terms of swirl was necessary.

In a last step it would be interesting to find out the results produced by a combination of the least leaking seal with the seal showing the most favorable rotordynamic response. In this case then the honeycomb seal is equipped with 8 swirl brakes on the circumference, as practiced with the see-through seal.

Figure 10 demonstrates that swirl brakes help to effectively reduce the cross-coupled stiffness and increase the direct damping of a honeycomb seal (TOR2938H). The exciting force of the honeycomb seal is now the same as for the see-through seal with smooth rotor and swirl brakes (TOR298DB).

Conclusions

The easy manageable and remarkably accurate identification procedure is suited for comparing labyrinth gas seal concepts with regard to the rotordynamic stability behavior. The investigations demonstrate the importance of considering not only the cross-coupled stiffness, but also the damping capability of the seal.

The use of swirl brakes at the entrance of the seal is highly effective. The swirl brakes reduce the preswirl and, as a consequence the cross-coupled stiffness, and they even increase the direct damping. In the investigated case only 8 swirl brakes on the circumference are sufficient to reduce the exciting force to a quarter of the initial value. Swirl brakes are perfectly well suited for a subsequent installation in turbomachines with stability problems.

Seals with labyrinth rotor and honeycomb stator cannot always be recommended due to the small direct damping. With high whirling frequencies the traditional concept with smooth rotor generates the smaller excitation. Nevertheless, the honeycomb seal considerably reduces the leakage rate. In conclusion, the use of a honeycomb seal in combination with swirl brakes turns out to be the optimum solution to minimize both the leakage and the rotordynamic excitation.

Nomenclature

a = Amplitude of the vibrational mode
C, E, K = Direct damping, inertia, stiffness coefficients
c, e, k = Cross damping, inertia, stiffness coefficients
c_{u0} = Entry swirl
F_T = Lateral force
Δr = Change in direct stiffness of the magnetic bearing
Δp = Pressure difference
Δq = Change in magnetic excitation at the stability limit
x, y = Coordinates of displacement
Ω = Whirling frequency

References

[1] Childs, D. W., 1993, *Turbomachinery Rotordynamics*, John Wiley and Sons, New York.

[2] Wagner, N. G., and Steff, K., 1996, ''Dynamic Labyrinth Coefficients From a High-Pressure Full-Scale Test Rig Using Magnetic Bearings,'' NASA Conference Publication to be published, proceedings of a workshop held at Texas A&M University, May 6–8.

[3] Childs, D. W., Nelson, C. E., Nicks, C., Scharrer, J., Elrod, D., and Hale, K., 1986, ''Theory Versus Experiment for the Rotordynamic Coefficients of Annular Gas Seals: Part 1—Test Facility and Apparatus,'' ASME J. Tribol., **108**, pp. 426–432.

[4] Hawkins, L., Childs, D., and Hale, K., 1989, ''Experimental Results for Labyrinth Gas Seals With Honeycomb Stators: Comparison to Smooth-Stator Seals and Theoretical Predictions,'' ASME J. Tribol., **111**, pp. 161–168.

[5] Yu, Z., and Childs, D., 1996, ''A Comparison of Rotordynamic Coefficients and Leakage Characteristics for Hole-Pattern Gas Damper Seals and Honeycomb Seal,'' NASA Conference publication to be published, proceedings of a workshop held at Texas A&M University, May 6–8.

[6] Wright, D. V., 1983, ''Labyrinth Seal Forces on a Whirling Rotor,'' ASME Applied Mechanics Division, *Proceedings of the Symposium on Rotor Dynamical Instability*, Vol. 55, M. L. Adams, Jr., ed., ASME, New York, pp. 19–31.

[7] Millsaps, K. T., and Martinez-Sanches, M., 1993, ''Rotordynamic Forces in Labyrinth Seals: Theory and Experiment,'' NASA Conference Publication 3239, proceedings of a workshop held at Texas A&M University, pp. 179–207.

[8] Matros, M., Neumer, T., and Nordmann, R., 1994, ''Identification of Rotordynamic Coefficients of Centrifugal Pump Components Using Magnetic Bearings,'' preprints of the Fifth International Symposium on Transport Phenomena and Dynamics of Rotating Machinery (ISROMAC-5), May 8–11, Kaanapali, HI, pp. 55–72.

[9] Wagner, N. G., and Pietruszka, W. D., 1988, ''Identification of Rotordynamic Parameters on a Test Stand with Magnetic Bearings,'' *Magnetic Bearings Proceedings of the First International Symposium*, ETH Zürich, G. Schweitzer, ed., Springer-Verlag, Berlin, pp. 289–299.

[10] Kwanka, K., and Mair, R., 1995, ''Identification of Gas Seal Dynamic Coefficients Based on the Stability Behavior of a Rotor,'' *Proceedings of the 1st European Conference of Turbomachinery—Fluid Dynamic and Thermodynamic Aspects*, Mar. 1–3, Erlangen-Nürnberg, Germany, VDI-Report 1186, pp. 297–309.

[11] Ulbrich, H., 1988, ''New Test Techniques Using Magnetic Bearings,'' *Magnetic Bearings Proceedings of the First International Symposium*, ETH Zürich, G. Schweitzer, ed., Springer-Verlag, Berlin, pp. 281–288.
[12] Kwanka, K., 1995, ''Variation of Fluid Flow Forces in Seals With Rotor Bending,'' DE-Vol. 84-2, *1995 Design Eng. Technical Conf.*, Vol. 3—Part B, DE-Vol. 84-2, ASME, New York, pp. 1277–1282.
[13] Leie, B., and Thomas, H.-J., 1980, ''Self-Excited Rotor Whirl Due to Tip-Seal Leakage Forces,'' NASA Conference Publication 2133, proceedings of a workshop held at Texas A&M University, pp. 303–316.
[14] Benckert, H., and Wachter, J., 1980, ''Flow Induced Spring Coefficients of Labyrinth Seals for Application in Rotordynamics,'' NASA Conference Publication 2133, proceedings of a workshop held at Texas A&M University, pp. 1–17.
[15] Childs, D., and Ramsey, C., 1991, ''Seal-Rotordynamic-Coefficient Test Results for a Model SSME ATD-HPFTP Turbine Interstage Seal With and Without a Swirl Brake,'' ASME J. Tribol., **113**, pp. 198–203.
[16] Kwanka, K., and Nagel, M., 1996, ''Experimental Rotordynamic Coefficients of Short Labyrinth Gas Seals,'' NASA Conference Publication to be published, proceedings of a workshop held at Texas A&M University, May 6–8.

M. P. Mignolet
Department of Mechanical
and Aerospace Engineering,
Arizona State University, Tempe, AZ 85287-6106

C.-C. Lin
Department of Mechanical Engineering,
Chungchou Institute of Technology,
Yuanlin, Changhua Hsien, Taiwan, R.O.C

B. H. LaBorde
Department of Mechanical
and Aerospace Engineering,
Arizona State University, Tempe, AZ 85287-6106

A Novel Limit Distribution for the Analysis of Randomly Mistuned Bladed Disks[1]

A recently introduced perturbation technique is employed to derive a novel closed form model for the probability density function of the resonant and near-resonant, steady state amplitude of blade response in randomly mistuned disks. In its most general form, this model is shown to involve six parameters but, in the important practical case of a pure stiffness (or frequency) mistuning, only three parameters are usually sufficient to completely specify this distribution. A series of numerical examples are presented that demonstrate the reliability of this three-parameter model in accurately predicting the entire probability density function of the amplitude of response, and in particular the large amplitude tail of this distribution, which is the most critical effect of mistuning.
[DOI: 10.1115/1.1339001]

Introduction

The high sensitivity of both the free and forced responses of bladed disks to mistuning, i.e., to the existence of small blade-to-blade differences in their structural properties that arise for example during the manufacturing process, has been known and studied for at least three decades (see in particular the works of Whitehead [1]; Ewins [2]; Huang [3]; Griffin and Hoosac [4]; Kielb and Kaza [5]; Kaza and Kielb [6]; Basu and Griffin [7]; Sinha [8]; Wei and Pierre [9,10]; Sinha and Chen [11]; Mignolet and Lin [12]; and references therein). It has in particular often been found that this phenomenon leads, in two ways, to a dramatic increase in the computational effort required for an accurate determination of the forced response of bladed disks. First, the presence of variations in the structural characteristics of the blades across the disk directly violates the assumption of periodicity, which is often employed to reduce the number of degrees-of-freedom required to model the bladed disk. Second, and more importantly, the unpredictability of the small changes in the structural parameters of the blades has quite naturally led to the modeling of certain of these properties as random variables. Then, the response of a given blade cannot be expressed by a single deterministic quantity; rather it is characterized by a probability density function from which all information, such as mean and variance of the response, probability that the response exceeds a certain threshold, etc., can easily be computed. The determination of these probability density functions from a similar description of the random structural properties of the blades is, however, not a simple task since the relationship between the response of a multi-degree-of-freedom system and its masses, stiffnesses, and damping coefficients is severely nonlinear, especially in the neighborhood of the resonances.

To circumvent this difficult problem, many investigators turned to Monte Carlo simulations. This technique requires first that a large number of sets of the structural properties of the blades be simulated consistently with their specified probabilistic model. Then, for each of these bladed disks, deterministic structural dynamics methods are employed to determine the corresponding response of the blades. Finally, a statistical analysis of the produced population of blade responses is conducted to obtain an estimate of the corresponding probability density function. Clearly, the Monte Carlo simulation technique is very versatile but it is also extremely computationally intensive; 100,000 simulations of bladed disks have sometimes been used (see Mignolet and Lin [12], in particular) to obtain accurate estimates of the tails of the probability density function. Although a novel, adaptive perturbation scheme has been introduced (Lin and Mignolet [13]) that leads to substantial savings in the computational effort associated with the Monte Carlo simulation approach, this technique still remains numerically intensive.

Alternate approaches to evaluate the distribution of the response of the blades have also been suggested. The combined closed-form perturbation (CFP) method developed by Mignolet and Lin [12] relies on an interpretation of the equations of motion of the bladed disk as a transformation, from one set of random variables to another, i.e., from the structural properties of the blades to their responses. Accordingly, a multifold integral representation of the probability density function of the amplitude of blade response has been derived which can be evaluated in the limit of a small mistuning. This technique provides a nonparametric description of the probability density function since this curve is computed point per point.

The advantage of nonparametric estimation techniques, such as the CFP method but also as the Monte Carlo simulation approach, is of course their capability to resolve accurately the details of the sought distribution. This is, however, often achieved at the expense of a large computational effort, as already demonstrated in connection with the Monte Carlo simulation strategy. This situation can be remedied by turning to parametric estimation techniques, i.e., by specifying a model for the distribution that includes only a few parameters, say from one to six, and evaluating these unknowns to satisfy at best some available data. The success of such approaches is clearly conditioned by the availability of a meaningful functional form of the distribution of blade amplitude.

The use of parametric techniques for the estimation of the probability density function of the amplitude of blade response has been the focus of some previous investigations. In particular, Sinha [8] and Sinha and Chen [11] relied on the central limit theorem to motivate the use of the normal distribution to approximate the probability density function of the response of the blades. As such, their approach required knowledge of only the first two moments, i.e., mean and variance, of the response of the blades; given these two quantities, all other moments and probabilities could easily be estimated. Unfortunately, a thorough parametric study (Mignolet and Lin [12]) has demonstrated that

[1]An early version of this paper has been presented at the Turbo Expo'96 in Birmingham, United Kingdom (ASME Paper 96-GT-414).

Contributed by the Structures & Dynamics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the S&D Division May 4, 2000; final revision received by the ASME Headquarters June 9, 2000. Editor: H. D. Nelson.

this Gaussian model does not represent very accurately the probability density function of the amplitude of blade response in structural dynamic systems that are representative of turbomachines. In this light, the goal of the present investigation is to partially remedy this situation by providing a new parametric model that accurately matches the distribution of the resonant and near-resonant response of the blades in a broad range of situations. The parameters of this model can, in particular, be estimated from the values of the first moments of the amplitude of blade which can be computed by relying on a small size simulation, as was done here for simplicity, or on moment estimation techniques as already accomplished with the present model by Mignolet and Hu [14]. Note finally that the present investigation focuses solely on the probabilistic aspects of the response of a typical blade, the corresponding analysis for the highest responding blade on the disk has been conducted elsewhere (Mignolet et al. [15]).

Since the derivation of the distribution model relies extensively on the adaptive perturbation scheme recently introduced (Lin and Mignolet [13]), the main results of this technique are first briefly reviewed.

The Adaptive Perturbation Scheme

Modeling the bladed disk as a linear N-degree-of-freedom system leads to the set of N equations of motion

$$M\ddot{\mathbf{X}}+C\dot{\mathbf{X}}+K\mathbf{X}=\mathbf{F}(t), \tag{1}$$

where M, C, and K denote, respectively, the $N\times N$ mass, damping, and stiffness matrices of the system while $\mathbf{F}(t)$ and $\mathbf{X}(t)$ represent the time-dependent excitation and response vectors the jth components of which will be written as $F_j(t)$ and $X_j(t)$, $j=1,2,\ldots,N$.

Mistuning of the bladed disk is easily modeled by allowing some or all of the components of the matrices M, C, and K to be random with standard deviations that are small with respect to their corresponding mean values. This situation has led quite naturally to the formulation of perturbation techniques among which is the adaptive perturbation scheme recently proposed (Lin and Mignolet [13]). In deriving this approach, it was first argued that the analysis of the N-degree-of-freedom system, Eq. (1), should be achieved not in the physical coordinates, $\mathbf{X}(t)$, as in Eq. (1), but rather in terms of a set of variables $\mathbf{q}(t)$ that accurately represent the contributions of the various mode shapes in the response. These new coordinates can be related to the physical ones by the relation

$$\mathbf{X}(t)=T\mathbf{q}(t), \tag{2}$$

where T denotes an $N\times N$ transformation matrix.

Selecting the columns of T to be the mode shapes of the mistuned bladed disk leads of course to a set of variables $\mathbf{q}(t)$ that correspond exactly with the system's modal coordinates. Unfortunately, this selection yields a *random* transformation matrix the probabilistic description of which is an extremely challenging task. Thus, it was suggested that the columns of T should be selected as the mode shapes of a fixed system that is ''close'' to the one described by Eq. (1). In view of the comprehensive qualitative analysis performed by Wei and Pierre [9,10], it was suggested (Lin and Mignolet [13]) to choose the columns of the transformation matrix T to be the mode shapes of the tuned system when the blade-to-blade coupling is large (with respect to the mistuning) and those of the decoupled system when this coupling is small (with respect to the mistuning).

Introducing the above-mentioned change of variables, Eq. (2), in Eq. (1) and premultiplying the resulting relations by T^T leads to the transformed equations of motion

$$M'\ddot{\mathbf{q}}+C'\dot{\mathbf{q}}+K'\mathbf{q}=\mathbf{F}'(t), \tag{3}$$

where the symbol $.^T$ denotes the operation of matrix transposition and

$$M'=T^TMT;\quad C'=T^TCT;\quad K'=T^TKT,\ \text{and}\ \mathbf{F}'=T^T\mathbf{F}.$$

When the excitation vector is a single frequency harmonic function of the form

$$F_j(t)=\mathrm{Re}[\overline{F}_j\ e^{i\omega t}]\quad j=1,2,\ldots,N, \tag{4}$$

where $\mathrm{Re}[Y]$ denotes the real part of an arbitrary complex number Y and $\overline{F}_j$ are constants, it is readily shown that the system reaches a steady state response of the form

$$q_j(t)=\mathrm{Re}[\overline{q}_j\ e^{i\omega t}]\quad j=1,2,\ldots,N. \tag{5}$$

Introducing Eqs. (4) and (5) into Eq. (1) leads to the linear system of complex equations

$$H\mathbf{Z}=\tilde{\mathbf{F}}, \tag{6}$$

where

$$\mathbf{Z}^T=[\overline{q}_1\overline{q}_2\overline{q}_3\cdots\overline{q}_N] \tag{7}$$

$$\tilde{\mathbf{F}}^T=[\overline{F}_1\overline{F}_2\overline{F}_3\cdots\overline{F}_N]. \tag{8}$$

Finally, the elements H_{jl} of the impedance matrix H can be expressed as

$$H_{jl}=K'_{jl}-\omega^2M'_{jl}+i\omega C'_{jl}, \tag{9}$$

where M'_{jl}, C'_{jl}, and K'_{jl} represent the jl elements of the transformed mass, damping, and stiffness matrices M', C', and K', respectively.

Considering the magnitude of the terms H_{jl}, it is first noted for the base configuration, which is the tuned system when the coupling is stronger than the mistuning and the decoupled system in the reverse case, that the matrix $H=\overline{H}$ is diagonal, i.e.,

$$\overline{H}_{jl}=0\qquad\text{for}\quad j\neq l. \tag{10}$$

Further, the magnitude of H_{jj} is ''small'' when the excitation frequency is close to one of the natural frequency ω_l of the base system but is otherwise ''large.'' On this basis, it is suggested to partition the response vector $\mathbf{q}(t)$ in terms of the components $\mathbf{q}_S(t)$ and $\mathbf{q}_L(t)$ which are associated with small and large diagonal elements of $\overline{H}$, respectively. That is,

$$\mathbf{q}(t)=\begin{bmatrix}\mathbf{q}_S(t)\\ \mathbf{q}_L(t)\end{bmatrix} \tag{11}$$

where $\mathbf{q}_S(t)$ and $\mathbf{q}_L(t)$ are vectors of dimensions $d\times 1$ and $(N-d)\times 1$, respectively. The above-mentioned partition can be achieved directly from the physical variables $\mathbf{X}(t)$ by selecting the transformation matrix T to be

$$T=[\Phi_S\Phi_L], \tag{12}$$

where Φ_S and Φ_L denote the $N\times d$ and $N\times(N-d)$ matrices whose columns are the base system's mode shapes associated with the natural frequencies ω_l that are close to (for Φ_S) and far from (for Φ_L) the excitation frequency ω.

Proceeding with this partitioning, it is found that the linear system of equations (6) becomes

$$\begin{bmatrix}H_{SS} & H_{SL}\\ H_{LS} & H_{LL}\end{bmatrix}\begin{bmatrix}\mathbf{Z}_S\\ \mathbf{Z}_L\end{bmatrix}=\begin{bmatrix}\tilde{\mathbf{F}}_S\\ \tilde{\mathbf{F}}_L\end{bmatrix}, \tag{13}$$

where the matrices H_{SS}, H_{SL}, H_{LS}, and H_{LL} are of respective dimensions $d\times d$, $d\times(N-d)$, $(N-d)\times d$, and $(N-d)\times(N-d)$ while the vectors $\mathbf{Z}_S$, $\mathbf{Z}_L$, $\tilde{\mathbf{F}}_S$, and $\tilde{\mathbf{F}}_L$ have d, $(N-d)$, d, and $(N-d)$ components.

The computation of the steady state components $\mathbf{Z}_S$ and $\mathbf{Z}_L$ can now be performed. Specifically, it is found that

$$\mathbf{Z}_S = G_{SS}[\tilde{\mathbf{F}}_S - H_{SL}H_{LL}^{-1}\tilde{\mathbf{F}}_L] \tag{14}$$

and

$$\mathbf{Z}_L = H_{LL}^{-1}[\tilde{\mathbf{F}}_L - H_{LS}\mathbf{Z}_S] \tag{15}$$

where

$$G_{SS} = [H_{SS} - H_{SL}H_{LL}^{-1}H_{LS}]^{-1}. \tag{16}$$

Resonant Limit Distribution (Large Coupling)

The determination of a limit distribution for the amplitude of blade response in the presence of a resonant or near-resonant excitation will be achieved by relying on the formulation of the adaptive perturbation scheme (Lin and Mignolet [13]) reviewed in the previous section. Specifically, in the presence of a large coupling, assume first that

(a) only one of the natural frequencies of the base system is close to the excitation frequency, i.e., $d=1$.

(b) The tuned bladed disk is a classically damped system with small damping ratios, as usually encountered in turbomachinery applications.

(c) The response of the bladed disk is dominated by the effects of the component of the excitation along the mode shape corresponding to the''close'' natural frequency. This assumption is quite natural for a lightly damped dynamic system; the components of the excitation along the other mode shapes are all off-resonant and thus will yield contributions to the response of the blades that are much smaller than the resonant one. Then, one can set $\tilde{\mathbf{F}}_L = 0$ in Eqs. (14) and (15).

Under these conditions, the steady state response of the bladed disk can be expressed as, see Eqs. (14) and (15),

$$Z_S = G_{SS}\tilde{F}_S \tag{17}$$

and

$$\mathbf{Z}_L = -H_{LL}^{-1}\mathbf{H}_{LS}Z_S = -G_{SS}H_{LL}^{-1}\mathbf{H}_{LS}\tilde{F}_S \tag{18}$$

in the modal coordinates, $\mathbf{q}(t)$, or equivalently, as

$$\begin{aligned} X_j(t) &= X_j^{(c)}\cos\omega t + X_j^{(s)}\sin\omega t = A_j\cos(\omega t - \varphi_j) = \mathbf{E}_j^T T\mathbf{q}(t) \\ &= \mathrm{Re}[\mathbf{E}_j^T[\Phi_S \Phi_L]\mathbf{Z}e^{i\omega t}] \\ &= \mathrm{Re}\{G_{SS}\tilde{F}_S\mathbf{E}_j^T[\Phi_S - \Phi_L H_{LL}^{-1}\mathbf{H}_{LS}]e^{i\omega t}\} \end{aligned} \tag{19}$$

in the physical coordinates. Note in the above expression that $\mathbf{E}_j$ designates the vector of dimension N whose components are all zero except the jth one which equals one.

To obtain simple expressions for the components of the response, $X_j^{(c)}$ and $X_j^{(s)}$, and the corresponding amplitude and phase, A_j and φ_j, it is further assumed that

(d) the level of mistuning is small enough to neglect the terms that are of second and higher order in the components of the deviation matrix $\delta H = H - \bar{H}$. This includes products such as $\delta H_{LL}\mathbf{H}_{LS}, \mathbf{H}_{SL}H_{LL}^{-1}\mathbf{H}_{LS}, \ldots,$

and finally that

(e) the imaginary part of $\bar{H}_{LL}^{-1}$ is small with respect to its real component. This assumption is quite natural since the damping level is small and the excitation frequency is not close to the natural frequencies associated with the mode shapes included in the matrix Φ_L. In such situations, the transfer function consists mainly of a real component.

Under these additional assumptions, the components of the response $X_j^{(c)}$ and $X_j^{(s)}$, and the corresponding amplitude and phase, A_j and φ_j, can be expressed as

$$X_j^{(c)} = \frac{(y_2y_1 + y_3y_0)\cos\phi_F - (y_2y_0 - y_3y_1)\sin\phi_F}{y_2^2 + y_3^2} \tag{20}$$

$$X_j^{(s)} = -\frac{(y_2y_1 + y_3y_0)\sin\phi_F + (y_2y_0 - y_3y_1)\cos\phi_F}{y_2^2 + y_3^2} \tag{21}$$

$$A_j^2 = \frac{y_0^2 + y_1^2}{y_2^2 + y_3^2} \tag{22}$$

and

$$\tan\varphi_j = \frac{X_j^{(s)}}{X_j^{(c)}} \tag{23}$$

where

$$\begin{aligned} y_0 &= \mathrm{Im}\{\mathbf{E}_j^T[\Phi_S - \Phi_L\bar{H}_{LL}^{-1}\mathbf{H}_{LS}]\}|\tilde{F}_S| \\ &= -\mathbf{E}_j^T\Phi_L\bar{H}_{LL}^{-1}\,\mathrm{Im}\{\mathbf{H}_{LS}\}|\tilde{F}_S| \end{aligned} \tag{24}$$

$$\begin{aligned} y_1 &= \mathrm{Re}\{\mathbf{E}_j^T[\Phi_S - \Phi_L\bar{H}_{LL}^{-1}\mathbf{H}_{LS}]\}|\tilde{F}_S| \\ &= \mathbf{E}_j^T[\Phi_S - \Phi_L\bar{H}_{LL}^{-1}\,\mathrm{Re}[\mathbf{H}_{LS}]]|\tilde{F}_S| \end{aligned} \tag{25}$$

$$y_2 = \mathrm{Re}[H_{SS}] \quad \text{and} \quad y_3 = \mathrm{Im}[H_{SS}] \tag{26}$$

with

$$\begin{bmatrix} H_{SS} \\ H_{LS} \end{bmatrix} = [\Phi_S\Phi_L]^T(K - \omega^2 M + i\omega C)\Phi_S \tag{27}$$

and

$$\cos\phi_F = \frac{\mathrm{Re}[\tilde{F}_S]}{|\tilde{F}_S|} \quad \text{and} \quad \sin\phi_F = \frac{\mathrm{Im}[\tilde{F}_S]}{|\tilde{F}_S|} \tag{28}$$

At this point, note that the variables y_0, y_1, y_2, and y_3 are linear combinations of the deviations in the components of the mass, damping, and stiffness matrices that are associated with the mistuning. Thus, under very general conditions (for example the Borel conditions, see Papoulis [16]), it can be argued that y_0, y_1, y_2, and y_3 are well approximated by Gaussian random variables with respective means μ_j and standard deviations σ_j, $j=0, 1, 2,$ or 3. In the absence of mistuning, the impedance matrix $H=\bar{H}$ is diagonal and the vector $\mathbf{H}_{LS}=\bar{\mathbf{H}}_{LS}$ vanishes identically. Thus, the variable y_0 is proportional to the mistuning terms and it is concluded that

$$\mu_0 = 0. \tag{29}$$

Considering next the term y_2, it is observed that the diagonal term $\bar{H}_{SS}$ of the tuned impedance matrix is purely imaginary at resonance so that the corresponding mean value μ_2 vanishes identically, i.e.,

$$\mu_2 = 0 \quad \text{at resonance.} \tag{30}$$

Stiffness mistuning. In the case of a pure stiffness mistuning, it is seen from Eq. (27) that the imaginary parts of $\mathbf{H}_{LS}$ and H_{SS} are, respectively, zero and a fixed constant. Then,

$$y_0 = 0 \tag{31}$$

and

$$y_3 = \text{deterministic constant} \tag{32}$$

so that the probabilistic model of the response given by Eqs. (20)–(22) can be rewritten in the form

$$X_j^{(c)} = \frac{(\bar{y}_2\cos\phi_F + \sin\phi_F)\bar{y}_1}{\bar{y}_2^2 + 1} \tag{33}$$

$$X_j^{(s)} = -\frac{(\bar{y}_2\sin\phi_F - \cos\phi_F)\bar{y}_1}{\bar{y}_2^2 + 1} \tag{34}$$

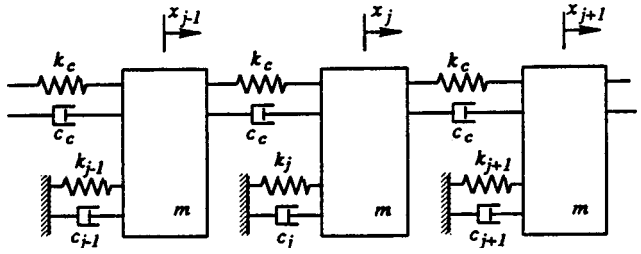


Fig. 1 One DOF per blade bladed disk model

$$A_j^2 = \frac{\overline{y}_1^2}{\overline{y}_2^2 + 1} \tag{35}$$

with

$$\overline{y}_1 = \frac{y_1}{y_3} \text{ and } \overline{y}_2 = \frac{y_2}{y_3}.$$

In the above model, $\overline{y}_1$ is a normal random variable with mean $\overline{\mu}_1$ and standard deviation $\overline{\sigma}_1$, $\overline{y}_2$ is a normal random variable with zero mean and standard deviation $\overline{\sigma}_2$.

Further, it has been shown in the case of the system displayed in Fig. 1 (see Lin [17]) that the random variables $\overline{y}_1$ and $\overline{y}_2$ are independent. Then, under these conditions, the probability density function of the amplitude A_j can be expressed as

$$p_A(a) = \frac{1}{\pi \overline{\sigma}_1 \overline{\sigma}_2} \int_1^\infty \frac{s^2}{\sqrt{s^2-1}} \exp\left[-\frac{s^2-1}{2\overline{\sigma}_2^2} \right] \times \left\{ \exp\left[-\frac{(as-\overline{\mu}_1)^2}{2\overline{\sigma}_1^2} \right] + \exp\left[-\frac{(as+\overline{\mu}_1)^2}{2\overline{\sigma}_1^2} \right] \right\} ds \tag{36}$$

and the corresponding first three moments of this distribution can be computed as

$$E[A_j] = \frac{\overline{\mu}_1}{\sqrt{2\pi}\overline{\sigma}_2} \exp\left[\frac{1}{4\overline{\sigma}_2^2} \right] K_0\left[\frac{1}{4\overline{\sigma}_2^2} \right] \tag{37}$$

$$E[A_j^2] = \sqrt{\frac{\pi}{2}} \frac{\overline{\sigma}_1^2 + \overline{\mu}_1^2}{\overline{\sigma}_2} \exp\left[\frac{1}{2\overline{\sigma}_2^2} \right] \text{erfc}\left[\frac{1}{\sqrt{2}\overline{\sigma}_2} \right] \tag{38}$$

$$E[A_j^3] = \frac{\overline{\mu}_1^3 + 3\overline{\mu}_1\overline{\sigma}_1^2}{2\sqrt{2\pi}\overline{\sigma}_2^3} \exp\left[\frac{1}{4\overline{\sigma}_2^2} \right] \left\{ K_1\left[\frac{1}{4\overline{\sigma}_2^2} \right] - K_0\left[\frac{1}{4\overline{\sigma}_2^2} \right] \right\}, \tag{39}$$

where K_0 and K_1 are the modified Bessel's functions of order 0 and 1 and erfc is the complimentary error function.

Given the three parameters $\overline{\mu}_1$, $\overline{\sigma}_1$, and $\overline{\sigma}_2$, the distribution and the moments of the amplitude of response of the blades can be numerically evaluated from Eqs. (36) to (39). Alternatively, Eqs. (36)–(39) can be used to estimate the distribution of the blade amplitudes, $p_A(a)$, given only the first three moments of this function. This computation is accomplished by first determining the parameters $\overline{\mu}_1$, $\overline{\sigma}_1$, and $\overline{\sigma}_2$ that correspond to the given first three moments of the response according to Eqs. (37)–(39). Then, the probability density function of the amplitudes is obtained from Eq. (36). Note that the first moments of the amplitude of blade response can be computed by relying on a small size simulation, as was done here for simplicity, or on moment estimation techniques as was accomplished by Mignolet and Hu [14].

It should finally be noted that Gaussian distributed components, $X_j^{(c)}$ and $X_j^{(s)}$, can be represented by the present model, Eqs. (33)–(35), by neglecting the variability of $\overline{y}_2$, or equivalently by assuming that $\overline{\sigma}_2 \ll 1$. This condition requires, in view of Eq. (26), that $\text{Re}[H_{SS}] \ll \text{Im}[H_{SS}]$ for all possible mistuned systems. The latter inequality can be rewritten with Eq. (27) as $\sigma_k/k_j \ll \zeta_j$, where k_j and ζ_j are the modal stiffness and damping ratio and σ_k is the standard deviation of the random stiffnesses. In general, however, σ_k/k_j and ζ_j are both of the same order of magnitude (typically a few percent) so that the response components $X_j^{(c)}$ and $X_j^{(s)}$ cannot be assumed to be Gaussian random variables.

Resonant Limit Distribution (Small Coupling)

In the presence of a small blade-to-blade coupling, it has been shown (see, for example, Wei and Pierre [9,10] and Lin and Mignolet [13]) that perturbation techniques yield reliable estimates of the steady state response of the blades provided that the base configuration is the uncoupled mistuned system and that the ''small quantities'' are the blade-to-blade coupling terms. Then, the modal matrix $[\Phi_S \Phi_L]$ reduces to the identity matrix and the coordinates $q_j(t)$ and $X_j(t)$ coincide. Further, selecting $d=0$ yields the following expression for the steady state response

$$\mathbf{Z} = (\hat{H} + \Delta H)^{-1}\tilde{\mathbf{F}} = (I_N + \hat{H}^{-1}\Delta H)^{-1}\hat{H}^{-1}\tilde{\mathbf{F}} = \hat{H}^{-1}\tilde{\mathbf{F}} - (\hat{H}^{-1}\Delta H)\hat{H}^{-1}\tilde{\mathbf{F}} + (\hat{H}^{-1}\Delta H)^2\hat{H}^{-1}\tilde{\mathbf{F}} + \ldots \tag{40}$$

where $\hat{H}$ and ΔH denote the impedance matrix of the uncoupled bladed disk and the matrix of the small coupling terms, respectively. Further, I_N designates the unit matrix of dimension $N \times N$. Considering the system shown in Fig. 1, it is readily shown that $\hat{H}$ is diagonal so that the response of the jth degree-of-freedom can be written as

$$X_j(t) = X_j^{(c)} \cos \omega t + X_j^{(s)} \sin \omega t = A_j \cos(\omega t - \varphi_j) \tag{41}$$

where

$$X_j^{(c)} = \frac{(y_2 y_1 + y_3 y_0)}{y_2^2 + y_3^2} \tag{42}$$

$$X_j^{(s)} = -\frac{(y_2 y_0 - y_3 y_1)}{y_2^2 + y_3^2} \tag{43}$$

$$A_j^2 = \frac{y_0^2 + y_1^2}{y_2^2 + y_3^2} \tag{44}$$

and

$$\tan \varphi_j = \frac{X_j^{(s)}}{X_j^{(c)}} \tag{45}$$

where

$$y_0 = \text{Im}\{\mathbf{E}_j^T[\tilde{\mathbf{F}} - (\Delta H \hat{H}^{-1})\tilde{\mathbf{F}} + (\Delta H \hat{H}^{-1})^2\tilde{\mathbf{F}} + \ldots]\} \tag{46}$$

$$y_1 = \text{Re}\{\mathbf{E}_j^T[\tilde{\mathbf{F}} - (\Delta H \hat{H}^{-1})\tilde{\mathbf{F}} + (\Delta H \hat{H}^{-1})^2\tilde{\mathbf{F}} + \ldots]\} \tag{47}$$

$$y_2 = k_j - m_j\omega^2 \quad \text{and} \quad y_3 = c_j\omega \tag{48}$$

with m_j, c_j, and k_j denoting the mass, damping coefficient, and stiffness associated with the jth blade. It can be shown that the model given by Eqs. (42)–(44) is still, approximately, applicable when each blade is represented by a multi-degree-of-freedom model provided that m_j, c_j, and k_j designate the jth blade's modal mass, damping coefficient, and stiffness corresponding to the natural frequency being excited, i.e., $\omega_l \approx \omega$.

Note the similarities between the models corresponding to small and large blade-to-blade coupling, i.e., Eqs. (42)–(45) and (20)–(23), respectively. In fact, Eqs. (20)–(23) reduce exactly to Eqs. (42)–(45) if the phase angle of the excitation ϕ_F is set to zero. Further, Eq. (30), and in the case of pure stiffness mistuning Eq. (32), also apply in the small coupling case. On the basis of these similarities, it is suggested that the model given by Eqs. (20)–(23) is an accurate representation of the random response of mistuned bladed disks in a broad range of situations.

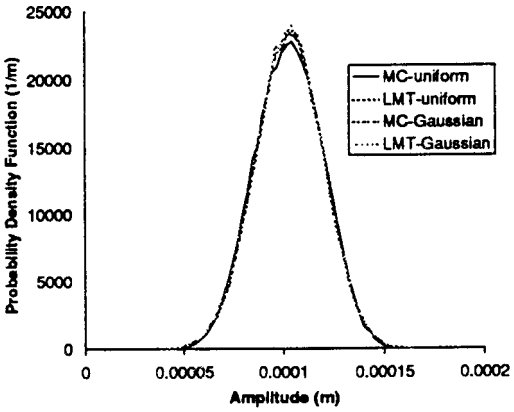


Fig. 2 Probability density functions of blade response, Gaussian and uniform distributions of stiffnesses, $N=24$, $r=3$, large coupling case. Monte Carlo (MC) simulations and limit distribution (LMT), Eq. (36).

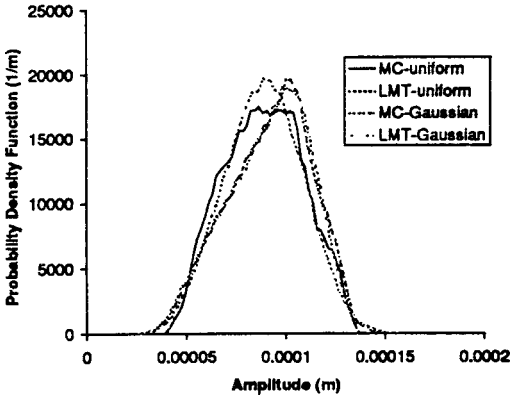


Fig. 3 Probability density functions of blade response, Gaussian and uniform distributions of stiffnesses, $N=24$, $r=3$, small coupling case. Monte Carlo (MC) simulations and limit distribution (LMT), Eq. (36).

Numerical Examples

To exemplify the usefulness of the limit distribution derived in this paper, the bladed disk shown in Fig. 1 is first considered. Following previous investigations (Sinha [8]; Sinha and Chen [11]; Mignolet and Lin [12]), the blades will be assumed to have the same mass ($m=0.0114$ kg), damping coefficient ($c_j=1.443$ Ns/m, approximately 1 percent of the critical value), but have different stiffnesses. Specifically, the parameters k_j are selected to be random variables with mean $k_t=430{,}000$ N/m and standard deviation $\sigma_k=8000$ N/m. Further, the aerodynamic and structural coupling between blades are modeled by springs and dashpots of common values k_C and c_C. In the sequel, it will be assumed that the coefficient c_C vanishes and that $k_C=45{,}430$ or 2271 N/m in the large and small coupling analyses, respectively.

It is further assumed that the blades are subjected to a resonant rth engine order excitation of the form of Eq. (4) with

$$\overline{F}_j=F_0e^{i\psi_j} \quad j=1,2,\ldots,N \tag{49}$$

$$\omega=\omega_r=\sqrt{\frac{k_t+4k_C\sin^2\left(\frac{\pi r}{N}\right)}{m}} \tag{50}$$

and

$$\psi_j=\frac{2\pi}{N}r(j-1). \tag{51}$$

In all numerical examples presented in the following, the value $F_0=1$ N has been used.

The assessment of the reliability of the model given by Eqs. (33)–(39) was accomplished by comparing the probability density functions of the amplitude of blade response predicted by this model and obtained by Monte Carlo simulation. More specifically, the three parameters $\overline{\mu}_1$, $\overline{\sigma}_1$, and $\overline{\sigma}_2$ were selected so that the first, second, and third moments of the random amplitude of blade response estimated by simulation exactly matched their model counterparts, Eqs. (37)–(39). Then, the probability density functions obtained from Eq. (36) and by simulation were plotted and a visual comparison of these two curves was accomplished. In some cases, small fluctuations in the obtained values of the parameters $\overline{\mu}_1$, $\overline{\sigma}_1$, and $\overline{\sigma}_2$ were achieved to obtain an improved matching between model and simulation distributions. Note that the simulation results correspond to the exact steady state responses of the blades, not to the simplified equations obtained under assumptions (a)–(e).

The large coupling limit was investigated first. Shown in Fig. 2 are the probability density functions of the amplitude of blade response corresponding to a third engine order ($r=3$) excitation for stiffness mistuning of both Gaussian and uniform distributions. Note, as predicted by the above-mentioned theoretical developments, that these two probability density functions are almost identical to each other, thereby demonstrating the existence of a ''limit distribution'' for the amplitude of blade response.

Also shown in Fig. 2 are the corresponding three-parameter approximations obtained from Eq. (36) with the parameters $\overline{\mu}_1$, $\overline{\sigma}_1$, and $\overline{\sigma}_2$ estimated as discussed previously. The excellent matching of these curves with the Monte Carlo results clearly demonstrates the validity of the three-parameter distribution of Eqs. (35) and (36). Additional results displaying a similar matching for different engine orders r, different numbers of blades N, and a near-resonant excitation, not presented here for brevity (see Lin [17]), confirm the reliability of the three-parameter model in the large coupling case.

Turning now to the small coupling limit, $k_C=2271$ N/m, shown in Fig. 3 are the probability density functions obtained by Monte Carlo simulations for both Gaussian and uniform stiffness mistuning together with their respective three-parameter approximations. It should first be noted from this figure that the amplitude of blade response does not admit, in contrast to the large coupling case, a limit-type distribution. That observation is consistent with the definition of y_2, see Eq. (48). Further, as expected, the three-parameter model of Eqs. (35) and (36) provides a very accurate representation of the amplitude of response in the presence of

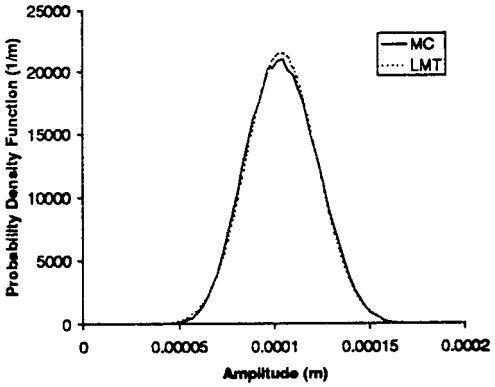


Fig. 4 Probability density functions of blade response, Gaussian distribution of stiffnesses, $N=24$, $r=0$, large coupling case. Monte Carlo (MC) simulations and limit distribution (LMT), Eq. (36).

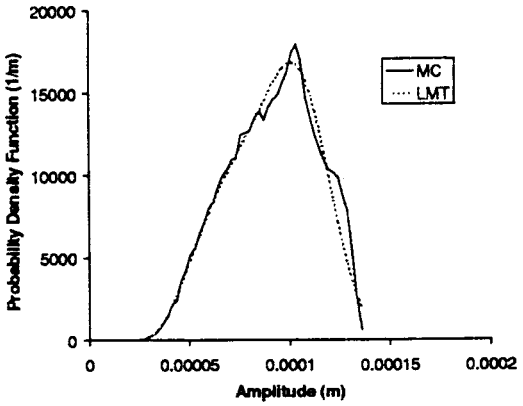


Fig. 5 Probability density functions of blade response, Gaussian distribution of stiffnesses, $N=24$, $r=0$, small coupling case. Monte Carlo (MC) simulations and limit distribution (LMT), Eq. (36).

Gaussian stiffnesses but is also quite reliable, especially in the important high amplitude tail, in the uniform case.

The probability density functions of the amplitude of blade response corresponding to the engine orders $r=0$ and $r=N/4=6$ were found to exhibit unusual features. In the large coupling limit and for $r=0$, this statistical distribution is very similar to its $r=3$ counterpart, see Figs. 2 and 4, but the response to a sixth engine order is characterized by much smaller values of the standard deviation of the blade vibration and of the estimated parameter $\overline{\sigma_2}$ than for other values of r. In fact, in this latter case a Gaussian approximation, i.e., Eqs. (35) and (36) with $\overline{\sigma_2}=0$, of the probability density function of the blade amplitude matches quite well the simulation results. In the small coupling limit, the case $r=N/4=6$ does not exhibit the unusual properties it has when k_C is large but it is the zeroth engine order excitation that distinguishes itself from other values of r. Nevertheless, the corresponding three-parameter approximation is quite reliable as can be seen from Fig. 5.

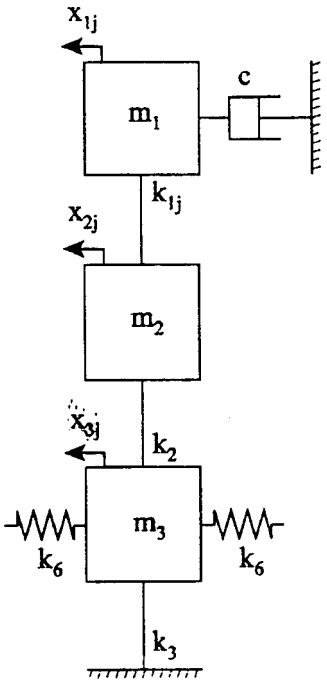


Fig. 6 Three DOF per blade bladed disk model

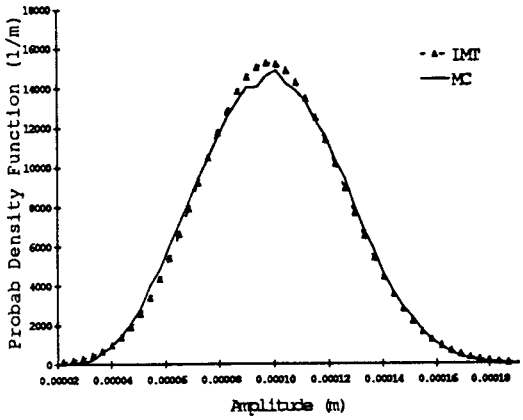


Fig. 7 Probability density function of response of DOF 1, Gaussian distribution of stiffnesses, $N=72$, $r=9$, 3 DOF model. Monte Carlo (MC) simulation and limit distribution (LMT), Eq. (36).

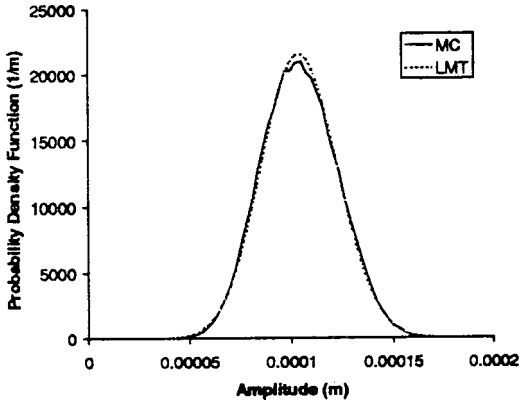


Fig. 8 Probability density function of blade response, Gaussian distribution of stiffnesses, nonconforming case, $c=0.7215$ Ns/m, small coupling case. Monte Carlo (MC) simulations and limit distribution (LMT), Eq. (36).

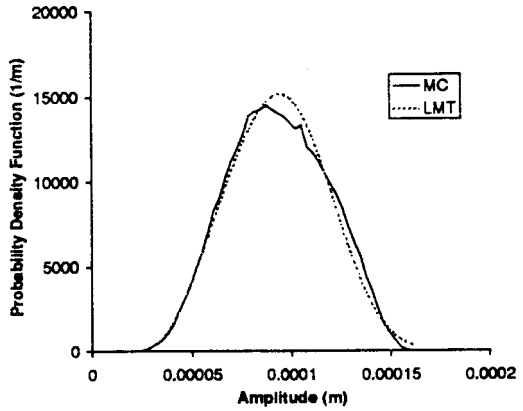


Fig. 9 Probability density function of blade response, Gaussian distribution of stiffnesses, nonconforming case, worst matching in the weak-to-strong coupling transition. Monte Carlo (MC) simulations and limit distribution (LMT), Eq. (36).

To further assess the accuracy of the limit distribution, Eq. (36), in predicting the statistical distribution of the amplitude of blade response, the three-degree-of-freedom per blade model considered by Griffin and Hoosac [4] and shown in Fig. 6 has been considered. The disk was assumed to support 72 blades the properties of which are (see Griffin and Hoosac [4])

$$m_1 = 0.0114\ \text{kg}; \quad m_2 = 0.0427\ \text{kg}; \quad m_3 = 0.0299\ \text{kg}$$

$$k_1(\text{mean}) = 430{,}300\ \text{N/m}; \quad k_2 = 17{,}350{,}000\ \text{N/m};$$

$k_3 = 7{,}521{,}000$ N/m; $k_6 = 30{,}840{,}000$ N/m; $c = 1.3814$ Ns/m.

Mistuning was simulated by letting the stiffnesses k_{1j} be Gaussian random variables with the above mean and a standard deviation of 8000 N/m. Shown in Fig. 7 are the probability density functions of the amplitude of response of the blade (degree-of-freedom 1) corresponding to a ninth engine order and a unit excitation force obtained by both simulation and the limit distribution. In this case again, Eq. (36) leads to a very good approximation of the Monte Carlo results.

As a final assessment of the usefulness of Eq. (36) in the analysis of randomly mistuned bladed disks, the applicability of this limit distribution in situations that do not conform to the stated assumptions, i.e., large or small coupling and hypotheses (a)–(e), was investigated. In this respect, note first that the neglect of the term $\mathbf{H}_{SL}H_{LL}^{-1}\mathbf{H}_{LS}$ (assumption (d)) in comparison to H_{SS} in the expression for G_{SS}, Eq. (16), is only valid when the damping terms $c\omega_j$ are typically larger than the level of mistuning. In view of the excellent results obtained when these quantities are approximately equal (see Figs. 2 and 3 where $c\omega_j \approx 8000$ N/m), it was desired to investigate the situation in which the damping coefficient is reduced to half of its previous value, i.e., $c = 0.7215$ Ns/m in the model of Fig. 1, while maintaining the standard deviation of stiffness mistuning to be 8000 N/m. Although a very small degradation in the reliability of the three-parameter approximation could be observed, the matching between model and simulation probability density functions is still excellent, see, for example, Fig. 8 for $k_C = 2271$ N/m. Finally, the matching between simulation results and the three-parameter model was also investigated for values of the coupling stiffness in the weak-to-strong transition region, i.e., for $k_C \in [5000; 20{,}000]$ N/m for the model of Fig. 1. It was found that the least satisfactory matching occurs in a narrow zone of coupling stiffnesses, i.e., between $k_C = 4000$ N/m and $k_C = 10{,}000$ N/m. However, even in the worst case situation, at $k_C = 6814$ N/m, see Fig. 9, the three-parameter model still provides a fair approximation of the distribution of blade response.

Summary

The determination of a reliable model of the probability density function of the amplitude of the resonant and near-resonant response of the blades in mistuned bladed disks has been accomplished by relying on perturbation methods. The case of a strongly coupled bladed disk was first considered and a general representation of the response of the blades was derived, Eqs. (20)–(23), that involves six parameters. Interestingly, a separate treatment of the opposite case, i.e., of a weakly coupled bladed disk, has led to a similar model and thus demonstrated the generality of Eqs. (20)–(23).

The important case of a pure stiffness, or frequency, mistuning of a strongly coupled bladed disk was also considered. It was demonstrated in this situation that Eqs. (20)–(23) reduce to the three-parameter model given by Eqs. (33)–(35) for which a closed form expression of the probability density function was obtained, Eq. (36). The corresponding first three moments of the amplitude of response were also given and a parametric identification technique was described in which these moments are matched to their simulation counterparts. This straightforward approach has led to extremely reliable approximations of the Monte Carlo results for a wide variety of cases with different numbers of blades, engine orders, distributions of the random stiffnesses, etc.; see in particular Figs. 2–4 and 7. Although the reliability of the proposed three-parameter model is somewhat lessened in situations that do not conform to the stated assumptions, e.g., when the blade-to-blade coupling is in the strong-to-weak transition region, it still provides a good approximation of the important high amplitude tail of the distribution. The reliable matching between the simulation results and those predicted by Eq. (36) demonstrates the importance of the three-parameter model, Eqs. (33)–(35), and, in a broader sense, of the general six-parameter representation, Eqs. (20)–(23).

References

[1] Whitehead, D. S., 1966, ''Effect of Mistuning on the Vibration of Turbomachines Blades Induced by Wakes,'' J. Mech. Eng. Sci., **8**, pp. 15–21.

[2] Ewins, D. J., 1969, ''The Effects of Detuning Upon the Forced Vibrations of Bladed Disks,'' J. Sound Vib., **9**, pp. 65–79.

[3] Huang, W.-H., 1982, ''Vibration of Some Structures with Periodic Random Parameters,'' AIAA J., **20**, pp. 1001–1008.

[4] Griffin, J. H., and Hoosac, T. M., 1984, ''Model Development and Statistical Investigation of Turbine Blade Mistuning,'' ASME J. Vib., Acoust., Stress, Reliab. Des., **106**, pp. 204–210.

[5] Kielb, R. E., and Kaza, K. R. V., 1984, ''Effects of Structural Coupling on Mistuned Cascade Flutter and Response,'' ASME J. Eng. Gas Turbines Power, **106**, pp. 17–24.

[6] Kaza, K. R. V., and Kielb, R. E., 1985, ''Vibration and Flutter of Mistuned Bladed-Disk Assemblies,'' J. Propul., **1**, pp. 336–344.

[7] Basu, P., and Griffin, J. H., 1986, ''The Effects of Limiting Aerodynamic and Structural Coupling in Models of Mistuned Bladed Disk Vibration,'' ASME J. Vib., Acoust., Stress, Reliab. Des., **108**, pp. 132–139.

[8] Sinha, A., 1986, ''Calculating the Statistics of Forced Response of a Mistuned Bladed Disk Assembly,'' AIAA J., **24**, pp. 1797–1801.

[9] Wei, S.-T., and Pierre, C., 1988, ''Localization Phenomena in Mistuned Assemblies with Cyclic Symmetry. Part I. Free Vibrations,'' ASME J. Vib., Acoust., Stress, Reliab. Des., **110**, pp. 429–438.

[10] Wei, S.-T., and Pierre, C., 1988, ''Localization Phenomena in Mistuned Assemblies with Cyclic Symmetry. Part II. Forced Vibrations,'' ASME J. Vib., Acoust., Stress, Reliab. Des., **110**, pp. 439–449.

[11] Sinha, A., and Chen, S., 1989, ''A Higher Order Technique to Compute the Statistics of Forced Response of a Mistuned Bladed Disk,'' J. Sound Vib., **130**, pp. 207–221.

[12] Mignolet, M. P., and Lin, C. C., 1993, ''The Combined Closed Form-Perturbation Approach to the Analysis of Mistuned Bladed Disks,'' ASME J. Turbomach., **115**, pp. 771–780.

[13] Lin, C. C., and Mignolet, M. P., 1997, ''An Adaptive Perturbation Scheme for the Analysis of Mistuned Bladed Disks,'' ASME J. Eng. Gas Turbines Power, **119**, pp. 153–160.

[14] Mignolet, M. P., and Hu, W., 1998, ''Direct Prediction of the Effects of Mistuning on the Forced Response of Bladed Disks,'' ASME J. Eng. Gas Turbines Power, **120**, pp. 626–634.

[15] Mignolet, M. P., Hu, W., and Jadic, I., 2000, ''On the Forced Response of Harmonically and Partially Mistuned Bladed Disks. II. Partial Mistuning and Applications,'' Int. J. Rotating Mach., **6**, pp. 43–56.

[16] Papoulis, A., 1984, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York.

[17] Lin, C. C., 1994, ''Forced Response of Mistuned Bladed Disks: Identification and Prediction,'' Ph.D. Dissertation, Arizona State University.

Identification of Mistuning Characteristics of Bladed Disks From Free Response Data—Part I

M. P. Mignolet

A. J. Rivas-Guerra

Department of Mechanical and Aerospace Engineering, Arizona State University, Tempe, AZ 85287-6106

J. P. Delor

Vehicle Dynamics and Mechanisms, General Motors Corporation, 2000 Centerpoint Parkway, Pontiac, MI 48341-3147

The focus of the present two-part investigation is on the estimation of the dynamic properties, i.e., masses, stiffnesses, natural frequencies, mode shapes and their statistical distributions, of turbomachine blades to be used in the accurate prediction of the forced response of mistuned bladed disks. As input to this process, it is assumed that the lowest natural frequencies of the blades alone have been experimentally measured, for example, in a broach block test. Since the number of measurements is always less than the number of unknowns, this problem is indeterminate in nature. In this first part of the investigation, two distinct approaches will be investigated to resolve the shortfall of data. The first one relies on the imposition of as many constraints as needed to ensure a unique solution to this identification problem. Specifically, the mode shapes and modal masses of the blades are set to their design/tuned counterparts while the modal stiffnesses are varied from blade to blade to match the measured natural frequencies. The second approach, based on the maximum likelihood principle, yields estimates of all the structural parameters of the blades through the minimization of a specified ''cost function.'' The accuracy of these two techniques in predicting the forced response of mistuned bladed disks will be assessed on simple dynamic models of the blades. [DOI: 10.1115/1.1338949]

Introduction

A large number of investigations have focused, in the last thirty years, on the assessment of the effects of mistuning, i.e., small blade-to-blade variations of their mechanical and/or geometrical properties, on both the free and forced responses of the entire disk. Most notably, it was shown that the free response can exhibit mode shapes which are localized to one or a few blades in contrast to their tuned counterparts which extend to the entire disk [1]. Further, it was demonstrated that the amplitude of the forced response varies greatly within a given disk from one blade to another, sometimes by a factor of 1 to 4 or larger, even when the relative fluctuations in the blade properties are as low as 1 to 2 percent (see [2–5], and references therein). Not only have these phenomenological observations been clarified but computational approaches have been developed to quantitatively assess these effects [6–13].

In most of these analyses, the variations of the blade properties were either specified or modeled as random variables with given means, standard deviations, and probability density functions, thereby generally sidestepping the issue of *characterization* of the mistuning from experimental data. In order to have a fully realistic description of the mistuning phenomenon, it would however be necessary to accurately estimate the fluctuations in the model parameters, i.e., stiffness, mass and damping terms, the correlations between these properties, and possibly their statistical distributions. The goal of the present work is thus the estimation of the statistical properties of these dynamic parameters by relying on experimental data.

A previous investigation of this identification problem [14] has relied on the availability of measurements of the forced response of consecutive blades in a series of disks, as obtained for example from strain gauges or light probes. The methodology proposed therein quite successfully recovered not only the blade structural properties but also the somewhat elusive blade-to-blade interaction terms which are known to play an important role in the analysis of mistuning. However, the measurement of the forced response of a series of bladed disks, which is required to obtain a good characterization (means, standard deviations, correlations, probability density functions, . . .) of the variations of the structural properties of the blades, is a time-demanding and expensive task. Thus, it would be highly desirable to develop an alternative methodology for the determination of the mistuned properties of the blades that utilizes the results of simple and inexpensive dynamic tests. In this light, the goal of the present two-part paper is more specifically to assess the usefulness of blade alone free response (broach block) test data to accomplish this task. Certainly, such measurements do not allow the determination of the blade-to-blade interaction effects, in contrast to the procedure recently introduced by Mignolet and Lin [14]. However, they can be used, as will be demonstrated, to provide structural dynamic models that exhibit forced resonance response characteristics, e.g., mean and largest amplitudes, that very closely approximate those corresponding to the tested bladed disks given the blade-to-blade interaction effects. Before addressing the specific details of the identification problem of concern here, it is important to first assess the feasibility of this task.

Free and Forced Response Indetermination

The characteristics of the free response of a system, i.e., natural frequencies and mode shapes, do not constitute sufficient information to recover uniquely its entire dynamic model, i.e., its mass (M) and stiffness (K) matrices. To accomplish this task, it is also necessary to dispose of either the modal masses or stiffnesses of the system. In this light, the identification of the mistuning characteristics of blades from only the measured values of the frequencies, as obtained in a broach block test, appears to be a severely indeterminate problem. Fortunately, it is the forced response of the bladed disk which is of concern here and it should be noted that the resonant response of a system does not depend either on all the details of the mass and stiffness matrices but rather only on the corresponding modal characteristics. This observation indicates that a reliable prediction of the forced response of mistuned bladed disks does not necessarily require the accurate determination of all their structural dynamic characteristics. Thus, it is pos-

Contributed by the Structures & Dynamics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the S&D Division, May 4, 1999; final revision received by the ASME Headquarters June 9, 1999. Editor: H. D. Nelson.

sible that an identification strategy be devised that leads to a very good estimation of the forced response while requiring only the knowledge of the natural frequencies. This property is especially true if close approximations of the mass and stiffness matrices of the system are available, as is the case in mistuned bladed disks for which the tuned structural properties can be estimated accurately, by finite elements for example.

To exemplify this situation, consider a single-degree-of-freedom system whose mass (m') and stiffness (k') deviate only slightly from their assumed values m and k. Then, the amplitude of steady state response to an excitation of frequency ω and magnitude F_0 is

$$A = \frac{F_0/m'}{\sqrt{(\omega'^2-\omega^2)^2+4\zeta^2\omega^2\omega'^2}}, \tag{1}$$

where $\omega' \approx \sqrt{k'/m'}$ is the measured natural frequency of the system and ζ its damping ratio. It is seen from the above-mentioned relation that approximating the correct mass m' by its assumed value m leads only to an error of order $|m'-m|/m' \ll 1$. However, replacing both m' and k' by m and k implies the approximation $\omega' \approx \sqrt{k/m}$, which may lead to a serious error in the estimation of the amplitude of response when $\omega \approx \omega'$ (resonant condition) and $\zeta \ll 1$ (small damping ratio) which are of interest here. In this case, the availability of the natural frequency is a necessary but also sufficient condition for the reliable estimation of the amplitude of the forced resonant response.

Identification Algorithms Formulation

Random Modal Stiffnesses (RMS) Approach. The above discussion suggests a first identification strategy of mistuned bladed disks in which the mass matrix is taken equal to its tuned counterpart while the mistuning affects only the stiffness matrix. An additional assumption is necessary if each blade is modeled as an N-degree-of-freedom system with $N>1$. Indeed, in this case there are at least $N(N+1)/2$ stiffness elements to be evaluated but only as many as N measured natural frequencies ω_n, $n=1,2,\ldots,N$. In the spirit of the above discussion, it is then assumed that the mode shapes of the blades are unaffected by mistuning so that the modal matrix Φ is given by its tuned counterpart Φ_t. If the mode shapes are normalized with respect to the mass matrix, i.e., if

$$\Phi_t^T M_t \Phi_t = I_N, \tag{2}$$

where I_N denotes the $N \times N$ identity matrix, it is found that the stiffness matrix of a blade is readily obtained as

$$K_{\mathrm{RMS}} = \Phi_t^{-T} \Lambda \Phi_t^{-1} \tag{3}$$

where Λ is the diagonal matrix containing the measured natural frequencies, i.e.,

$$\Lambda = \mathrm{diag}(\omega_1^2, \omega_2^2, \ldots, \omega_N^2). \tag{4}$$

According to the above strategy, only the modal stiffnesses of the blades are affected by mistuning and thus only these quantities vary randomly from blade to blade.

Maximum Likelihood (ML) Method. In the random modal stiffnesses strategy described in the previous section, the indetermination in the estimation of the mass and stiffness matrices is resolved by adding enough constraints (constancy of mode shapes and mass matrix) to yield a unique solution. A different approach can be undertaken in which the unknown structural parameters are chosen to optimize a certain ''cost function.''

To formulate such an optimization problem, note first that small fluctuations of the structural parameters of the blades do, intuitively, appear more likely to occur than large ones, or more generally that the possible variations in these parameters are not all equally ''plausible.'' Thus, when facing several distinct sets of structural parameters yielding natural frequencies equal to their measured values, it is suggested here to select the one which had the highest probability of occurrence. This selection strategy is consistent with the maximum likelihood principle (see Benjamin and Cornell [15], for a complete presentation) which is based on the fact that the outcome of an experiment that has the maximum probability is the most likely one to be observed. Reversing the argument, it is assumed in the maximum likelihood formulation that the event that occurred was indeed the most probable one. Then, it is assumed here that *the ''true'' values of the masses and stiffnesses of the model of a blade are the most likely ones given the observed values of the natural frequencies of that blade*.

The application of this estimation strategy to the present problem can be described as follows. Due to nonzero tolerances, in-service wear, and/or small variations in material properties, the stiffness and mass matrices of a blade, K and M, are not exactly equal to their design counterparts, K_t and M_t, but rather fluctuate around these quantities. Then, introduce first the vector $\mathbf{X}$ that contains the unknown deviations of the elements of the stiffness and mass matrices K and M from their tuned values. Next, according to the maximum likelihood principle, it is suggested to seek the values of $\mathbf{X}$ that maximize the joint probability function, $p_{\mathbf{X}}(\mathbf{X})$, of the deviations of the elements of the stiffness and mass matrices of the blades. This function must be maximized under the constraints that the dynamic model of a given blade does indeed possess the corresponding measured natural frequencies ω_j, $j=1,2,\ldots,m$. Note that the number of such measurements does not necessarily equal the number of degree-of-freedom, i.e., $m \leqslant N$. These conditions can mathematically be expressed as

$$C_j \equiv \det(K - M\omega_j^2) = 0 \quad j=1,2,\ldots,m. \tag{5}$$

The above procedure clearly relies on a prior knowledge of the probability density function $p_{\mathbf{X}}(\mathbf{X})$ of the stiffness and mass elements of the blade model but in practical situations this information is not available. Fortunately, it has recently been demonstrated [11] that the forced response of mistuned bladed disks appears to depend only slightly, in most cases, on the shape of the distribution of the mistuned blade properties; it depends mainly on the variances and correlations of these variables. Thus, one can assume a specific shape of the unknown probability density function $p_{\mathbf{X}}(\mathbf{X})$ without seriously compromising the accuracy of the prediction of the forced response analysis. On this basis and on the common place character of the normal distribution in physical processes, it is assumed that the stiffnesses and masses are jointly normally distributed so that

$$p_{\mathbf{X}}(\mathbf{X}) = \frac{1}{(2\pi)^{s/2}\sqrt{\det(K_{\mathbf{XX}})}} \exp\left[-\frac{1}{2}\mathbf{X}^T K_{\mathbf{XX}}^{-1}\mathbf{X}\right], \tag{6}$$

where s is the number of stiffness and mass parameters to estimate, and $K_{\mathbf{XX}}$ denotes the $s \times s$ unknown covariance matrix of the deviations of these variables with respect to their means. The goal of the maximum likelihood approach, which is to estimate the vector $\mathbf{X}$ that maximizes $p_{\mathbf{X}}(\mathbf{X})$, Eq. (6), can be accomplished by selecting the vector $\mathbf{X}$ that minimizes the argument of the exponential, i.e.,

$$\epsilon = \mathbf{X}^T K_{\mathbf{XX}}^{-1} \mathbf{X}. \tag{7}$$

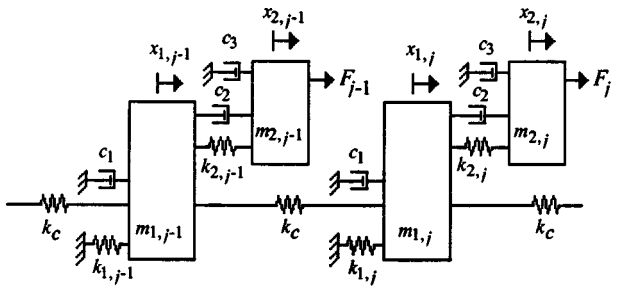
In addition to yielding the measured natural frequencies, see Eq. (5), it will be assumed that the blade model must also satisfy the following two sets of linear constraints:

$$D_j \equiv \mathbf{d}_j^T \mathbf{X} = e_j \quad j=1,2,\ldots,p, \tag{8}$$

and

$$\tilde{D}_j \equiv \tilde{\mathbf{d}}_j^T \mathbf{X} = \tilde{e}_j \quad j=1,2,\ldots,q. \tag{9}$$

The first set of constraints is introduced to account for possible relationships existing between the elements M_{ij} and K_{ij} of the mass and stiffness matrices as a consequence of the selected dy-



$$F_j = F_o \cos\left(\omega_r t + \frac{2\pi r(j-1)}{N_b}\right) \qquad N_b = 24 \text{ blades} \quad F_o = 1 \text{ N}$$

$$\sigma_{m_1}/m_{1t} = \sigma_{m_2}/m_{2t} = \sigma_{k_1}/k_{1t} = \sigma_{k_2}/k_{2t} = 0.005\,.$$

Model 1: $m_{1t} = 0.0780$ kg $m_{2t} = 0.0114$ kg $k_{1t} = 7{,}166{,}028$ N/m
$k_{2t} = 415{,}557$ N/m $k_c = 5{,}236{,}000$ N/m
$c_3 = 1.443$ Ns/m $c_1 = c_2 = 0$

Model 2: $m_{1t} = 0.0114$ kg $m_{2t} = 0.001055$ kg $k_{1t} = 430{,}000$ N/m
$k_{2t} = 36{,}411.44$ N/m $\zeta_1 = \zeta_2 = 1\%$ $k_c = 45{,}430$N/m

Fig. 1 Two-degree-of-freedom per blade disk models

namic model of the blades and holds the same for all blades analyzed. For example, in the spring–mass system shown in Fig. 1, it can be shown that $K_{22} = -K_{12}\ (=k_2)$, or $K_{12}+K_{22}=0$, and $M_{12}=0$. On this basis, the set of Eq. (8) will be referred to as the modeling constraints.

The second set of constraints, Eq. (9), differs from the first one, Eq. (8), in that the vectors $\tilde{\mathbf{d}}_j$ and/or the coefficients $\tilde{e}_j$ are not the same for all blades considered, as were $\mathbf{d_j}$ and e_j, but rather vary from blade to blade. Such a set of constraints is encountered in particular when the dynamic model of the blade is required to match not only measured natural frequencies ω_j but also observed mode shapes $\boldsymbol{\phi}_j$. Indeed, in this case, it is necessary to also enforce the conditions $(K-\omega_j^2 M)\boldsymbol{\phi}_j = 0$ which are linear in the elements of the stiffness and mass matrices K and M and thus can be written in the form of Eq. (9). Further, the corresponding parameters $\tilde{\mathbf{d}}_j$ and $\tilde{e}_j$ will involve the measured values of ω_j and $\boldsymbol{\phi}_j$, and thus will vary from blade to blade.

The separation of the modeling and mode shape constraints into two separate sets, Eqs. (8) and (9), may appear unnecessary at first since they are both of the same form. Note, however, that the existence of the *deterministic* constraints, Eq. (8), implies that the random deviations of the masses and stiffnesses from their tuned values are not linearly independent or equivalently that a subset $\mathbf{X}_2$ of p of these structural parameters can be expressed as a linear combination of the remaining ones stored in the vector $\mathbf{X}_1$. Further, this linear deterministic dependence of $\mathbf{X}_2$ on $\mathbf{X}_1$ is associated (see Lutes and Sarkani [16]) with a singular covariance matrix $K_{\mathbf{XX}}$ so that the definition of ϵ, Eq. (7), appears unclear in view of the lack of existence of the inverse $K_{\mathbf{XX}}^{-1}$. In fact, this situation is not unexpected since the linear deterministic dependence of $\mathbf{X}_2$ on $\mathbf{X}_1$ implies that the statistical distribution of $\mathbf{X}_2$ is completely dictated by its counterpart for $\mathbf{X}_1$. Correspondingly, the most likely vector $\mathbf{X}$ will be found by maximizing the likelihood of $\mathbf{X}_1$ alone. Equivalently, the proposed mistuning identification approach will rely on the minimization of

$$\epsilon_1 = \mathbf{X}_1^T K_{\mathbf{X}_1\mathbf{X}_1}^{-1} \mathbf{X}_1\,, \tag{10}$$

where $K_{\mathbf{X}_1\mathbf{X}_1}$ is the covariance matrix of $\mathbf{X}_1$ and is obtained by deleting the rows and columns of $K_{\mathbf{XX}}$ corresponding to the elements of $\mathbf{X}$ that belong to the dependent subset $\mathbf{X}_2$.

To complete the separation of the identification problem into independent ($\mathbf{X}_1$) and dependent ($\mathbf{X}_2$) subsets of structural parameters, note that the partitioning of $\mathbf{X}$ into $\mathbf{X}_1$ and $\mathbf{X}_2$ can be reversed by writing

$$\mathbf{X} = L_1^T \mathbf{X}_1 + L_2^T \mathbf{X}_2\,, \tag{11}$$

where the matrices L_1 and L_2 contain only zeros and ones.

The minimization of ϵ_1, under the constraints given by Eqs. (5), (8), and (9), will be accomplished by the Lagrange multipliers technique. Specifically, the new function to minimize is

$$f = \mathbf{X}_1^T K_{\mathbf{X}_1\mathbf{X}_1}^{-1} \mathbf{X}_1 + \sum_{j=1}^{m} \lambda_j C_j + \sum_{j=1}^{p} \mu_j D_j + \sum_{j=1}^{q} \tilde{\mu}_j \tilde{D}_j\,, \tag{12}$$

where m, p, and q are the number of measured natural frequencies, modeling constraints, and observed mode shapes, respectively, and λ_j, μ_j, and $\tilde{\mu}_j$ are the corresponding Lagrange multipliers associated with each set of constraints. The minimization of f for each blade will involve its partial differentiation with respect to each element of $\mathbf{X}_1$ and $\mathbf{X}_2$ and each Lagrange multiplier. Then, setting these partial derivatives to zero provides a sufficient number of equations to estimate all the elements of the vectors $\mathbf{X}_1$ and $\mathbf{X}_2$, i.e., the masses and stiffnesses of the blade considered.

When the blade structural model is a N-degree-of-freedom system, the natural frequency constraints given by Eq. (5) are nonlinear conditions, i.e., polynomials of order N, in the elements of the stiffness and mass matrices. This nonlinearity renders a direct determination of the minimum of f a difficult problem. To palliate this situation, it is proposed to proceed with a linearization of the frequency constraints. This approximation can in fact be viewed as the first step in an iterative solution of the exact minimization of f, Eq. (12). Specifically, a first-order Taylor expansion of Eq. (5) around estimates K' and M' of the stiffness and mass matrices yields (see Rivas-Guerra [17] for a proof)

$$1 + \text{trace}[(\Delta K - \omega_j^2 \Delta M)(K' - \omega_j^2 M')^{-1}] = 0 \tag{13}$$

for $j = 1,2,\ldots,m$ where $\Delta K = K - K'$ and $\Delta M = M - M'$. Expanding the trace in the above equation leads to the following linearized form of this constraint as

$$C_j \equiv 1 + \mathbf{a}_j^T (\mathbf{X} - \mathbf{X}') = 0 \tag{14}$$

where the vector $\mathbf{a}_j$ depends on the current estimates K' and M' of the stiffness and mass matrices, or equivalently $\mathbf{a}_j = \mathbf{a}_j(\mathbf{X}')$.

At this point, the minimization of Eq. (12) can be performed with the linearized constraints, Eq. (14). Specifically, it is found that the deviations in structural properties stored in the vectors $\mathbf{X}_1$ and $\mathbf{X}_2$ satisfy the symmetric system of equations

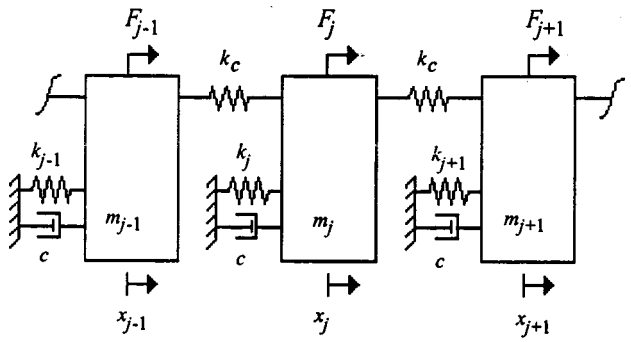
$$\begin{bmatrix} 2K_{\mathbf{X}_1\mathbf{X}_1}^{-1} & [L_1\mathbf{a}] & [L_1\mathbf{d}] & [L_1\tilde{\mathbf{d}}] & 0 \\ [L_1\mathbf{a}]^T & 0 & 0 & 0 & [L_2\mathbf{a}]^T \\ [L_1\mathbf{d}]^T & 0 & 0 & 0 & [L_2\mathbf{d}]^T \\ [L_1\tilde{\mathbf{d}}]^T & 0 & 0 & 0 & [L_2\tilde{\mathbf{d}}]^T \\ 0 & [L_2\mathbf{a}] & [L_2\mathbf{d}] & [L_2\tilde{\mathbf{d}}] & 0 \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 \\ \boldsymbol{\lambda} \\ \boldsymbol{\mu} \\ \tilde{\boldsymbol{\mu}} \\ \mathbf{X}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{g} \\ \mathbf{e} \\ \tilde{\mathbf{e}} \\ \mathbf{0} \end{bmatrix} \tag{15}$$

where $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$, $\tilde{\boldsymbol{\mu}}$, $\mathbf{e}$, and $\tilde{\mathbf{e}}$ are the vectors containing the values of λ_j, μ_j, $\tilde{\mu}_j$, e_j, and $\tilde{e}_j$, respectively. Further, the elements g_j of the vector $\mathbf{g}$ are defined as

$$g_j = -1 + \mathbf{a}_j^T L_1^T \mathbf{X}_1' + \mathbf{a}_j^T L_2^T \mathbf{X}_2' \quad \text{for} \quad j = 1,2,\ldots,m. \tag{16}$$

Finally, the notation $[L_1\mathbf{a}]$ defines the matrix $[L_1\mathbf{a}_1\ L_1\mathbf{a}_2 \cdots L_1\mathbf{a}_m]$ and similarly for $[L_1\mathbf{d}]$ and $[L_1\tilde{\mathbf{d}}]$.

It was found in general that the solution of Eq. (15) with the frequency constraints linearized with respect to the tuned system, i.e., $\mathbf{X}_1' = \mathbf{X}_2' = 0$, yields a blade model whose natural frequencies match very well their measured counterparts, ω_j, $j = 1,2,\ldots,m$, except when two or more of these values are very



$$F_j = F_o \cos\left(\omega_r t + \frac{2\pi r(j-1)}{N_b}\right)$$

m_t=0.0114 kg k_t=430,000 N/m c=1.443 Ns/m

k_c=45,430 N/m N_b=24 blades F_o = 1 N

$$\sigma_k = 0.01k_t \text{ and } \sigma_m = 0.01\sqrt{2}\, m_t$$

Fig. 2 Single-degree-of-freedom per blade disk model

close to each other. In these near veering conditions, the minimization of Eq. (12) with the nonlinear frequency constraints was achieved iteratively by solving Eq. (15) first with $\mathbf{X}_1' = \mathbf{X}_2' = 0$. Then, the corresponding solution vectors $\mathbf{X}_1$ and $\mathbf{X}_2$ were used to update the estimates K' and M' of the stiffness and mass matrices in Eq. (13), or equivalently the values of $\mathbf{X}_1'$ and $\mathbf{X}_2'$, and the process was repeated until the natural frequencies of the blade model matched their measured counterparts to an acceptable level of accuracy (see [18] for more details).

Two important observations can be drawn from the above developments. First, the random model stiffness (RMS) approach can be viewed as a special case of the constrained maximum likelihood (ML) method in which the masses and mode shapes have been fixed to their tuned counterparts by using Eqs. (8) and (9), respectively. Second, note that the ML formulation developed above is equivalent to the constrained minimization of a weighted sum of the square deviations of the structural parameters from their tuned values. Thus, given a set of natural frequencies, the maximum likelihood estimates of the structural parameters always correspond to the masses and stiffnesses closest, in the above sense, to their tuned counterparts that agree with these frequencies. Certainly, the corresponding dynamic model of the blade is the most likely one to be present but there exists a nonzero probability for larger excursions of the structural parameters from their tuned values. It is thus concluded that the ML method will in general underpredict somewhat the level of mistuning. Since the RMS formulation is a special case of the ML approach, it can be expected that it will also lead to an underprediction of the mistuning.

Numerical Results

To assess the validity of the procedures developed in the previous section, a series of numerical experiments were performed in which the structural properties, i.e., masses and stiffnesses, of a large number (240,000) of blades were simulated as random variables. For simplicity, each blade was modeled either as a one- or as a two-degree-of-freedom dynamic system, see Figs. 1 and 2. Note that the nominal values of the parameters of the latter model (Fig. 2) have already been used in prior investigations of the mistuning phenomenon [6,8] while the tuned values of the two-degree-of-freedom system (Fig. 1, model 1) were selected to match at best the three-degree-of-freedom model considered by Griffin and Hoosac [2] . Once the ensemble of masses and stiffnesses (''original'' values) were generated, the natural frequencies of each blade were determined and served as input to the RMS

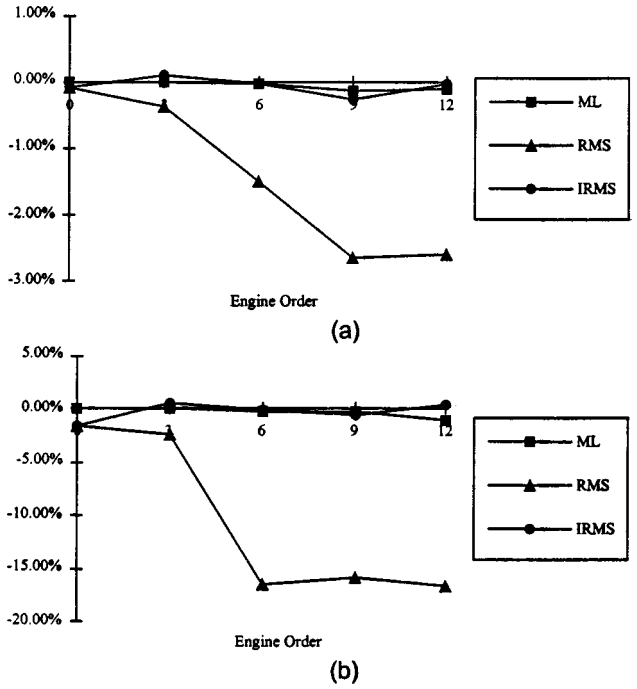


Fig. 3 Errors (*a*) in the mean and (*b*) in the standard deviation of the maximum amplitude of resonant response of the blade. ML, RMS, and IRMS approaches, single-degree-of-freedom blade model excited at the tuned natural frequency corresponding to the *r*th engine order, ω_r^t.

and ML procedures to recover estimates of these structural properties. Following the earlier discussion on the indetermination of this identification problem, it is not expected that the masses and stiffnesses recovered by either the RMS or the ML approach would match their original counterparts. However, it is desired that the characteristics of the forced response, i.e., mean, standard deviation, and distribution of the amplitudes of vibrations, of disks supporting blades with the ''recovered'' structural properties be close to their counterparts for blades with the ''original'' masses and stiffnesses. Thus, the ensuing numerical results will focus almost exclusively on comparing the forced response predictions obtained from both the ''original'' and ''recovered'' structural properties of the blades for the engine order excitations shown in Figs. 1 and 2.

Effects of Mass and Blade Alone Mode Shape Variations. One of the most salient differences in the formulations of the random modal stiffnesses (RMS) and maximum likelihood (ML) methods is the absence in the former approach of any mass and/or mode shape variability, although blade-to-blade fluctuations of these structural properties are expected to be present in physical disks. To assess the adequacy of this assumption, it is proposed here to first proceed with a thorough parametric study and then to seek theoretical confirmation by analyzing the simplest mistuned system, i.e., the uniformly mistuned bladed disk in which all the blades are identical but differ from the design/tuned ones.

Mass Mistuning: Importance. As a starting point to quantify the reliability of the modeling of mistuning by the random modal stiffnesses approach, consider the single-degree-of-freedom per blade model shown in Fig. 2 and assume that the masses and stiffnesses of different blades are independent of each other and normally distributed with standard deviations $\sigma_k = 0.01k_t$ and $\sigma_m = 0.01\sqrt{2}m_t$. Then, shown in Fig. 3 are the relative errors in the predicted mean and standard deviation of the amplitude of the highest responding blade for different engine order (r) excitations at the corresponding tuned natural frequency ω_r. Note that, as the

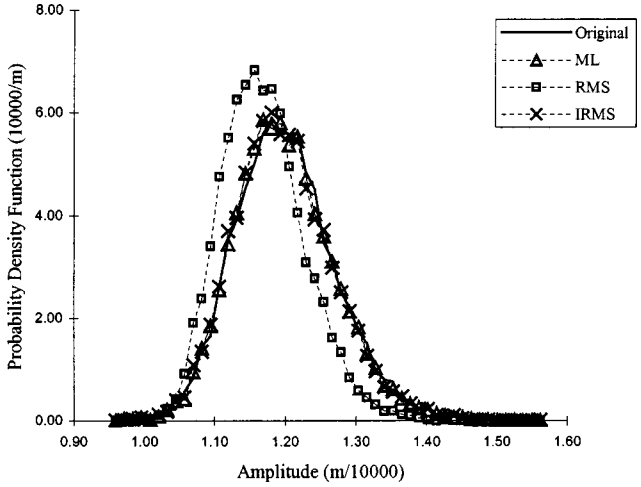


Fig. 4 Probability density function of the amplitude of the highest responding blade. Original parameters, estimated by ML, RMS, and IRMS approaches, single-degree-of-freedom blade model excited at the tuned natural frequency corresponding to the ninth engine order, ω_9^t.

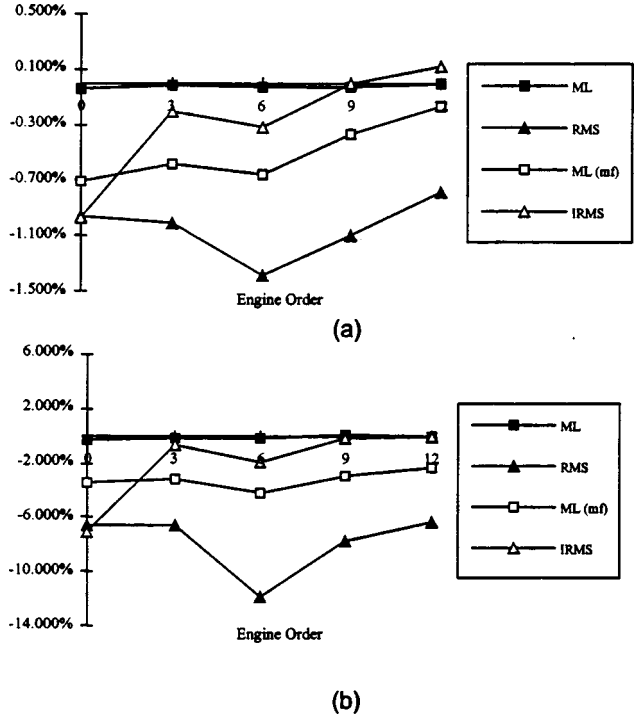


Fig. 5 Errors (*a*) in the mean and (*b*) in the standard deviation of the maximum response of the mass m_2. ML, ML with first mode fixed (mf), RMS and IRMS strategies, two-degree-of-freedom blade model 1 excited at the lowest tuned natural frequency corresponding to the *r*th engine order, $\omega_{r,1}^t$.

engine order increases, the errors in means and standard deviations also increase. In particular, it is seen that the errors in standard deviations can be as high as 20 percent for engine orders larger than 6.

Further, it should be emphasized that the errors in means and standard deviations for the blade exhibiting the maximum response are both negative, which suggests that the RMS approach could seriously underpredict the maximum amplitudes caused by mistuning. To confirm this expectation and assess the importance of the errors shown in Fig. 3, the probability density function of the amplitude of the highest responding blade was investigated for different values of r. Note first that for the zeroth engine order excitation ($r=0$), the original and the RMS curves almost exactly match (see [17]) which is expected since the corresponding errors in mean and standard deviation are very low, see Fig. 3. Further, for $r=5$, the probability density function of the blade exhibiting the maximum response, computed on the basis of the RMS structural properties shows a slight shift to the left with respect to its exact counterpart. For $r=6$, the magnitude of this shift is sharply amplified and remains almost constant for $r>6$, see Fig. 4 for $r=9$. This difference between the original and RMS probability density functions is clearly associated with the large errors in means and standard deviations shown in Fig. 3.

The effect of a shift to the left of the probability density function is to decrease the probability that the amplitude exceeds large values. An example of this situation is displayed in Table 1, where the probabilities that the maximum amplitude exceeds 1.3 times the tuned amplitude are presented for both the original set of structural properties and those estimated by the RMS approach. It is seen for low values of the engine order that the probabilities of exceeding this level are well predicted by estimating the structural properties with the RMS approach but for $r \geqslant 6$ this method fails to provide even a reasonable estimate of these probabilities. In fact, it is seen from Table 1 that the RMS estimation strategy yields unsafe estimates of the probabilities.

Table 1 Probabilities that the maximum amplitude exceeds 1.3 times the tuned response, RMS and ML approaches, single-degree-of-freedom blade model excited at the tuned natural frequencies ω_r^t.

Engine Order	Original	RMS	Relative Error	ML	Relative Error
$r=0$	0.156	0.150	-3.8 %	0.155	-0.3 %
$r=5$	0.063	0.038	-39.7 %	0.063	-0.1 %
$r=9$	0.180	0.067	-62.8 %	0.175	-2.8 %
$r=12$	0.239	0.106	-55.6 %	0.231	-3.3 %

To further confirm the above findings, the two-degree-of-freedom system shown in Fig. 1 (model 1) was also considered. Mistuning was simulated in the system by modeling the blade parameters (k_1, k_2, m_1, and m_2) as independent normal random variables, each one of which was characterized by a coefficient of variation of 0.5 percent, i.e., $\sigma_{m_1}/m_{1t}=\sigma_{m_2}/m_{2t}=\sigma_{k_1}/k_{1t}=\sigma_{k_2}/k_{2t}=0.005$. Note that this small amount of variability was found to be large enough to cause severe mistuning problems (see Griffin and Hoosac [2] for results on a similar model).

Then, shown in Fig. 5 are the relative errors in mean and standard deviation of the amplitude of the highest responding mass m_2 as computed from the RMS estimated structural parameters. Note that the disk was excited in the rth engine order mode in the first frequency band of the bladed disk. As in the one-degree-of-freedom model, it can be seen that the errors increase with the engine order but not monotonically. Further, the probability density functions of the maximum amplitudes of the response of the degree-of-freedom two, which is shown for $r=9$ in Fig. 6, exhibits a definite shift to the left as seen in connection with the one-degree-of-freedom blade model, Fig. 4. This figure suggests again that the probability that the amplitude exceeds a certain value could be seriously underpredicted by relying on structural parameters estimated by the RMS approach. To confirm this suspicion, presented in Table 2 is a comparison, original versus RMS, of the probabilities that the maximum amplitudes of the degrees-of-freedom one and two at the ninth engine order excitation exceed their exact mean values by one standard deviation. These results demonstrate again that the RMS strategy can lead to a serious

Table 2 Probabilities that the maximum amplitude exceeds its mean value by one standard deviation, RMS approach, two-degree-of-freedom blade model excited at the lowest tuned natural frequency corresponding to the ninth engine order, $\omega^t_{9,1}$.

D.O.F	Original	RMS	Relative Error
1	0.149	0.048	-67.8 %
2	0.147	0.068	-53.7 %

underestimation of the forced response of mistuned bladed disks thereby corroborating the findings of the one-degree-of-freedom blade model study.

Mass Mistuning: Uniformly Mistuned Bladed Disk. The reliability of the forced response estimates obtained through the RMS approach at the zeroth engine order is not unexpected, since this excitation emphasizes blade alone behavior which parallels the broach block testing strategy. In order to understand the problems encountered at high engine orders, note first that the natural frequencies of a tuned disk [19,12] are given by

$$\omega_r^2 = \frac{k_t + 4k_c \sin^2(\pi r/N_b)}{m_t} \tag{17}$$

for the one-degree-of-freedom per blade model shown in Fig. 2. In Eq. (17), N_b is the number of blades on the disk and k_c is the coupling stiffness (see Fig. 2). It is seen from Eq. (17) that each natural frequency of the disk involves the sum of the blade alone stiffness and a coupling term. When the disk is mistuned, the symmetry of the system is lost and Eq. (17) cannot in general be used. However, consider the rare but possible situation in which all the blades happen to have the same mass (m') and stiffness (k') which are not identical to the tuned values. The exact natural frequencies of this uniformly mistuned disk are

$$\omega_r'^2 = \frac{k' + 4k_c \sin^2(\pi r/N_b)}{m'}. \tag{18}$$

Next, assume that the RMS identification strategy is used to estimate the structural properties of the blades. In view of Eqs. (2)–(4), it is found that

$$m_{\mathrm{RMS}} = m_t \tag{19a}$$

$$k_{\mathrm{RMS}} = m_{\mathrm{RMS}}\omega^2 = m_t \frac{k'}{m'} \tag{19b}$$

so that the natural frequencies of a uniformly mistuned disk whose structural properties have been recovered by the RMS approach would be

$$(\omega_r^2)_{\mathrm{RMS}} = \frac{k'}{m'} + \frac{4k_c \sin^2(\pi r/N_b)}{m_t}. \tag{20}$$

Comparing Eqs. (18) and (20), it is seen that the RMS approach cannot account for the randomness in the system natural frequencies associated with the term k_c/m' which is approximated through the RMS procedure by the constant k_c/m_t. Neglecting high order terms in the variations of $m' - m_t$, the difference between the exact and estimated disk frequencies can thus be estimated as

$$(\omega_r)_{\mathrm{RMS}} - \omega_r' \approx \frac{2k_c}{\omega_r'} \frac{(m' - m_t)}{m_t^2} \sin^2(\pi r/N_b). \tag{21}$$

In fact, Eq. (21) can be viewed as an expression for the change in the rth tuned natural frequency of the disk that is obtained as a consequence of variations in the blade mass and stiffness *while keeping the blade alone natural frequency constant*, i.e., from Eq. (19) $k'/m' = k_{\mathrm{RMS}}/m_{\mathrm{RMS}}$. Then, the nonvanishing right-hand side of Eq. (21) indicates that the RMS approach does not account properly for the variations in the *disk* natural frequencies that occur due to the mistuning in both masses and stiffnesses. This observation suggests that the poor reliability of the RMS-based forced response prediction, see Fig. 3, may be a result of a slightly erroneous assessment of the closeness between the excitation frequency and the disk natural frequency of the resonating mode. To confirm this assumption, note that the resonant forced response of the two mistuned systems, exact (original) and estimated by the RMS, will be similar if the differences between the respective natural frequencies are typically much less than the corresponding half-power bandwidths or equivalently if

$$\frac{k_c \sin^2(\pi r/N_b)}{k_t + 4k_c \sin^2(\pi r/N_b)} \frac{\sigma_m}{m_t} \ll \zeta, \tag{22}$$

where ζ is the damping ratio in the rth mode of the bladed disk. The above formula confirms the findings of Table 1 and Fig. 3, i.e., for low engine orders the difference in the natural frequencies of the disk modes are small enough to be within the bandwidth of the corresponding peaks and no real difference in behavior is predicted. Further, as the engine order, r, increases the difference between frequencies also increases as a fraction of the bandwidth and the responses for the original bladed disk and for the one whose properties were estimated by the RMS method become quite different from each other so that large errors are observed.

The criterion given by Eq. (22) indicates that the coupling stiffness, k_c, and the damping ratio ζ should also play important roles in the reliability of the RMS approach. Specifically, an increase in the coupling stiffness and/or a decrease in the damping ratio should provoke a worsening of the RMS-based forced response prediction. A thorough study of the effects of these parameters (see Rivas-Guerra, [17]) has indeed confirmed these expectations obtaining in particular almost linear relationships between the errors in mean and standard deviations and the coupling stiffness k_c.

A final validation of the above explanation of the poor reliability of the RMS-based forced response prediction and of the usefulness of Eq. (20) will be provided in Part II in connection with the improved random modal stiffnesses (IRMS) approach.

Maximum Likelihood Results. The above discussion has clearly demonstrated the need to estimate both stiffness and mass properties of the blades, as is accomplished in particular by the ML approach. The assessment of the reliability of this technique in predicting the forced response will be accomplished as was done in connection with the RMS method by comparing the characteristics, i.e., statistics and probability density function, of the forced response of the systems shown in Figs. 1 and 2. In this light, note first that the maximum likelihood approach relies on the availability of the covariance matrix $K_{\mathbf{XX}}$, see Eqs. (6) and (15). In this first phase of the investigation, the exact covariance matrices will be used in all cases but the selection of this matrix and its influence on the reliability of the ensuing forced response prediction will be discussed in detail in Part II.

The corresponding results for the one-degree-of-freedom model (Fig. 2) are presented in Fig. 3 with their RMS counterparts. Clearly, the errors implied by the ML approach are much smaller than the ones obtained through the RMS technique. In fact, it was found that these errors were almost one order of magnitude smaller than the corresponding values of the RMS approach. Another basis for comparing these two techniques is the probability density function of the amplitude of the highest responding blade. The corresponding curves, shown in Fig. 4 for $r = 9$, clearly demonstrate that the ML estimation strategy, in contrast to the RMS approach, provides estimates of the structural properties of the blades which lead to a very accurate prediction of the forced response of mistuned bladed disks. As a final check, the probabilities that the maximum response exceeds 1.3 times the tuned value were computed on the basis of the structural parameters estimated by the ML approach. Comparing the ML values with their RMS counterparts, see Table 1, it is clear that the maximum likelihood

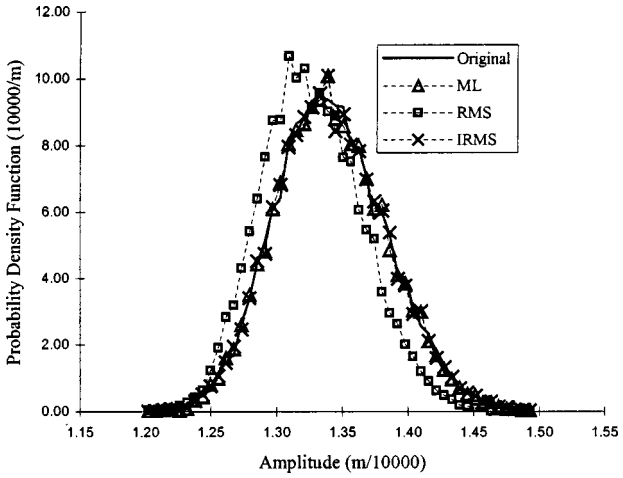


Fig. 6 Probability density function of the maximum amplitude of response of the mass m_2. Original parameters, estimated by ML, RMS, an IRMS approaches two-degree-of-freedom blade model 1 excited at the lowest tuned natural frequency corresponding to the ninth engine order, $\omega_{9,1}^t$.

(ML) approach performs consistently better than the random modal stiffness (RMS) technique for this first example.

Note from Fig. 3 that the ML errors increase with increasing engine order, as already observed in connection with the RMS approach. This behavior can be explained by the same arguments. Specifically, at low engine orders the response of the disk emphasizes blade alone behavior, which is what is observed experimentally in the broach block test. As the engine order is increased, the coupling between blades becomes more important and shifts the response behavior away from the test conditions. In other words, the parameters estimated by both the ML and the RMS methods yield the measured mistuned frequency of each blade, but they do not provide the correct mistuned frequencies of the corresponding bladed disks. Note, however, that the ML errors are much smaller than their RMS counterparts which increase more rapidly with the coupling stiffness. This increased accuracy of the ML method is clearly associated with the improved reliability of the approximation of the masses of the blades which are estimated to fit the measured natural frequencies at best as opposed to being set to the tuned value as in the RMS strategy.

The analysis of the forced response prediction for the two-degree-of-freedom per blade model shown in Fig. 1 (model 1) was carried out here by assuming that the two frequencies of the blades were measured, see Part II for a discussion of the effects of unmeasured frequencies. The results presented in Figs. 5 and 6 confirm the excellent reliability of the ML estimation procedure. Further, a comparison of the ML and RMS results demonstrates again the increased reliability of the former approach over the latter one as already observed in connection with the one-degree-of-freedom blade model.

Mode Shapes Mistuning: Importance. The previous discussion has emphasized in particular the importance of modeling the blade-to-blade mass (or modal masses) variations. It would be very desirable as well to assess the role of variations of the blade alone mode shapes on the reliability of the estimates of the forced response statistics of the corresponding mistuned bladed disks. To this end, the two-degree-of-freedom per blade model shown in Fig. 1 (model 1) was again considered but, in addition to the two frequency constraints used before (see Fig. 5), one constraint of the form of Eq. (9) was used with $\boldsymbol{\phi}_1$ equal to its tuned counterpart, $\boldsymbol{\phi}_1^t$, for all blades.

The relative errors in the mean and standard deviation of the maximum amplitude of response of mass 2 are presented in Fig. 5 (curve ML (mf), mf=mode fixed) together with their RMS and ML (no mode fixed) counterparts. It is clearly seen from these figures that the use of an identification methodology that does not include the variability of the mode shapes can lead to large errors in the subsequent prediction of the forced response of the corresponding mistuned bladed disks, especially at large engine orders.

Recognizing further that the difference between the RMS and mode constrained ML methods lies in the neglect, in the former approach, of the variability in the modal masses, provides a confirmation of the earlier finding that mass mistuning should indeed be estimated to obtain reliable forced response estimates.

Mode Shapes Mistuning: Uniformly Mistuned Bladed Disk. The analysis of uniformly mistuned bladed disks performed in connection with one-degree-of-freedom per blade models demonstrated a very important ''interaction'' between the blade-to-blade coupling and the mass mistuning, see Eq. (21). Further, this relation provided a solid basis to explain the steady worsening, with increasing engine order, of the forced response estimates obtained by the RMS approach. It is desired here to extend this analysis to the case of a multi-degree-of-freedom per blade model that exhibits mistuning in all properties, especially mode shapes, to assess the existence of a similar interaction but with the mode shape mistuning terms. To this end, note first that the rth engine order natural frequencies of the bladed disk, $\omega_{r,j}$, and the associated modal blade deflections, $\boldsymbol{\psi}_{r,j}$, satisfy the eigenvalue problem [19,12]

$$(K+K^{(r)})\boldsymbol{\psi}_{r,j}=\omega_{r,j}^2(M+M^{(r)})\boldsymbol{\psi}_{r,j}, \tag{23}$$

where $K^{(r)}$ and $M^{(r)}$ are $N\times N$ matrices that involve, respectively, the stiffness and mass coupling between blades as well as the engine order r (see Lin and Mignolet [12] for definitions). In the special case of a nearest-neighbors-only coupling through a stiffness matrix K_C, it is found that $K^{(r)}=4K_C\sin^2(\pi r/N)$ and $M^{(r)}=0$.

For the disk whose blade structural parameters have been estimated by the RMS approach, Eq. (23) takes the form

$$(K_{\mathrm{RMS}}+K^{(r)})\boldsymbol{\psi}_{r,j}^{(\mathrm{RMS})}=[\omega_{r,j}^{(\mathrm{RMS})}]^2(M_t+M^{(r)})\boldsymbol{\psi}_{r,j}^{(\mathrm{RMS})}. \tag{24}$$

At this point, recall that the dynamic model of the blades estimated by the RMS approach exhibits the correct natural frequencies so that

$$\Lambda=\Phi^TK\Phi=\Phi_t^TK_{\mathrm{RMS}}\Phi_t, \tag{25}$$

where, as in Eqs. (2)–(4), Φ and Φ_t denote the blade alone modal matrices of the mistuned and tuned blades normalized with respect to their corresponding mass matrices, M and M_t, and Λ is the diagonal matrix containing the measured natural frequencies. Then, premultiplying Eq. (23) by Φ^T and Eq. (24) by Φ_t^T lead to the two relations

$$(\Lambda+\Phi^TK^{(r)}\Phi)(\Phi^{-1}\boldsymbol{\psi}_{r,j})=\omega_{r,j}^2(I_N+\Phi^TM^{(r)}\Phi)(\Phi^{-1}\boldsymbol{\psi}_{r,j}) \tag{26}$$

and

$$(\Lambda+\Phi_t^TK^{(r)}\Phi_t)(\Phi_t^{-1}\boldsymbol{\psi}_{r,j}^{(\mathrm{RMS})})=[\omega_{r,j}^{(\mathrm{RMS})}]^2(I_N+\Phi_t^TM^{(r)}\Phi_t)\times(\Phi_t^{-1}\boldsymbol{\psi}_{r,j}^{(\mathrm{RMS})}). \tag{27}$$

Introducing the mode shape mistuning matrix $\delta\Phi=\Phi-\Phi_t$, it is found that

$$\Phi^TK^{(r)}\Phi=\Phi_t^TK^{(r)}\Phi_t+\delta\Phi^TK^{(r)}\Phi_t+\Phi_t^TK^{(r)}\delta\Phi+O(\delta\Phi^2) \tag{28}$$

and similarly for $M^{(r)}$. Introducing these approximations in Eq. (26) and proceeding to first order in the mistuning yields

$$\omega_{r,j}^2=[\omega_{r,j}^{(\mathrm{RMS})}]^2+2(\Phi_t^{-1}\boldsymbol{\psi}_{r,j}^{(\mathrm{RMS})})^T\times\Phi_t^T\{K^{(r)}-[\omega_{r,j}^{(\mathrm{RMS})}]^2M^{(r)}\}\delta\Phi(\Phi_t^{-1}\boldsymbol{\psi}_{r,j}^{(\mathrm{RMS})}). \tag{29}$$

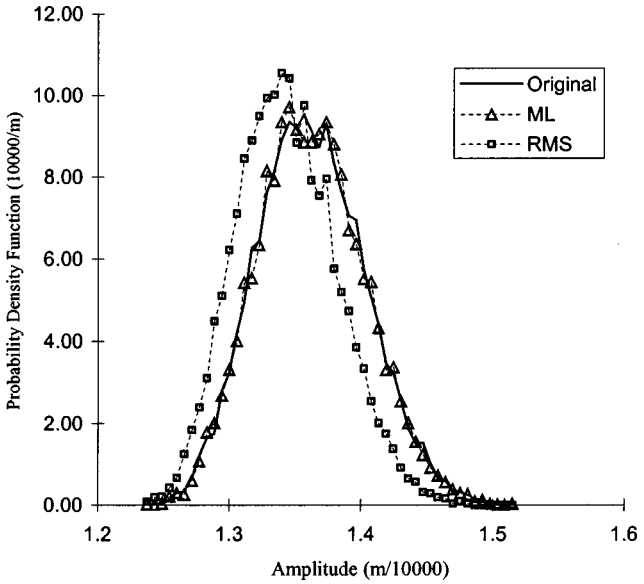


Fig. 7 Probability density function of the maximum amplitude of response of the mass m_2. Original parameters, estimated by ML and RMS approaches, two-degree-of-freedom blade model 1 excited in a ninth engine order excitation in a frequency sweep around the corresponding lowest tuned frequency, $\omega \in [0.95\omega^t_{9,1}, 1.05\omega^t_{9,1}]$.

Solving this relation for the RMS based natural frequencies $\omega^{(\mathrm{RMS})}_{r,j}$ and expressing the RMS modal blade deflection $\boldsymbol{\psi}^{(\mathrm{RMS})}_{r,j}$ as a perturbation of its tuned value $\boldsymbol{\psi}^t_{r,j}$ finally yields to first order

$$[\omega^{(\mathrm{RMS})}_{r,j}]^2 = \omega^2_{r,j} - 2[\boldsymbol{\psi}^t_{r,j}]^T [K^{(r)} - \omega^2_{r,j} M^{(r)}] \delta\Phi \Phi_t^{-1} \boldsymbol{\psi}^t_{r,j}. \tag{30}$$

As expected, it is found that the natural frequencies of the disk determined from the RMS estimates of the blade structural parameters deviate from their true values by a term that involves both the blade-to-blade coupling (in $K^{(r)}$ and $M^{(r)}$) and the mistuning in mode shape $\delta\Phi$. Note that the variations in masses are also present in Eq. (30) but do not explicitly appear because the modal matrices Φ and Φ_t have been normalized with respect to their corresponding mass matrices. In fact, in the case of a one-degree-of-freedom per blade model, one has $\Phi = 1/\sqrt{m}$ and $\Phi_t = 1/\sqrt{m_t}$ and Eq. (30) reduces to Eq. (21).

As before, the difference in the natural frequencies $\omega^{(\mathrm{RMS})}_{r,j}$ and $\omega_{r,j}$, see Eq. (30), exhibits many of the same features as the forced response of randomly mistuned bladed disks. Specifically, the presence of unmodeled mistuning in the mode shapes negatively affects the accuracy of the prediction of both the free and forced responses of the disk. Further, this effect is minimized when either the coupling is reduced or when the engine order is selected to be $r=0$ value at which the excitation emphasizes the blade alone response of the system.

Beside these important similarities in behavior, there are also some significant differences which must be highlighted. For example, it can be seen that the difference between $\omega^2_{r,j}$ and $[\omega^{(\mathrm{RMS})}_{r,j}]^2$ vanishes for $r=0$ but the errors in the predicted moments of the forced response, see Fig. 5, do not. More importantly, Eq. (30), as was Eq. (21), is indicative of a shift in natural frequencies of the bladed disk which only modifies the location of the resonance, not the magnitude of the corresponding forced response. If this observation also held in connection with a randomly mistuned disk, the distribution of the maximum amplitude of blade response observed on a given disk as the excitation frequency is varied would be the same if the parameters were the true ones, or estimated by either the RMS or ML method. The results of Fig. 7 and Table 3, however, contradict this scenario and thus provide a limitation to the validity of Eqs. (21) and (30) for the physical interpretation of the behavior of the RMS approach. Interestingly, note that the above-mentioned results are very similar to those relating to the maximum amplitude of response that occurs on a given disk at the resonance frequency of the tuned system, see Fig. 6 and Table 2, thereby demonstrating a strong consistency in the characteristics of the RMS approach.

Table 3 Probabilities that the maximum amplitude exceeds its mean value by one standard deviation, ML and RMS approaches, two-degree-of-freedom blade model 1 excited in a ninth engine order excitation in a frequency sweep around the corresponding lowest tuned frequency, $\omega \in [0.95\omega^t_{9,1}, 1.05\omega^t_{9,1}]$.

DOF	Original	RMS	ML	Error RMS	Error ML
1	0.1466	0.0443	0.1427	-69.77 %	-2.64 %
2	0.1476	0.0589	0.1476	-60.11 %	0.029 %

Summary

This investigation has focused on the estimation of the blade-to-blade variations of their structural properties, in particular masses and stiffnesses, from measurements of the natural frequencies of these blades only. Practically, this identification task does not represent an end result but rather a necessary first step in the accurate prediction of the forced response of turbomachine disks that would support the blades tested. It is thus in terms of forced response accuracy that the reliability of the suggested identification strategies has been assessed.

Since the knowledge of the natural frequencies of a system does not represent enough information to reconstitute its entire dynamic model, i.e., mass and stiffness matrices, the present identification problem is in fact indeterminate. In this first part of the investigation, two distinct procedures have been introduced to resolve the indeterminacy, the first one of which relies on the specification of additional constraints to ensure a unique solution to the problem. In the context of small mistuning, it was suggested that the mode shapes of the blades and the modal masses (or, in fact, equivalently the mass matrix) could be maintained equal to their design/tuned counterparts while the modal stiffnesses would vary from blade to blade to match the measured natural frequencies.

Clearly, this first procedure, termed the random modal stiffnesses (RMS) approach, is computationally quite attractive. Unfortunately, it was demonstrated that the reliability of the forced response of the corresponding mistuned disks predicted by relying on these estimates of the dynamic properties of the blades varies substantially, from excellent to poor, depending in particular on the level of blade-to-blade coupling, excitation characteristics (engine order number), etc. In all cases investigated, it was found that the RMS-based forced response prediction *underestimates* its exact counterpart leading to an unsafe estimate of the fatigue life. As an example of poor accuracy, it was demonstrated that the probability that the maximum amplitude of blade response exceeds specific levels can be underestimated by 50 percent or more! (see Tables 1–3).

The poor reliability of this technique was justified by investigating the natural frequencies, not of the blade alone, but of the entire bladed disk and noting that the dependence of the latter values on the masses and stiffnesses of the blades is different than it is for the former ones. Equivalently, given the mass distribution of the blades, their modal stiffnesses can be selected to provide either accurate estimates of the blade alone natural frequencies or of the natural frequencies of the entire bladed disk, but not of both. Thus, by matching precisely the measured values of the blade alone frequencies, the RMS approach leads to slightly erroneous estimates of the bladed disk modal characteristics, see Eq.

(21), which in turn provoke large discrepancies in the prediction of the forced response of the entire system near or at resonance in the presence of a small damping ratio.

To palliate this situation, a second strategy was introduced that aims at estimating all of the structural parameters of the system. Then, to resolve the indeterminacy associated with the shortfall of measurements, it was suggested to select the dynamic parameters of the blades to be the ''most probable ones given the observed values of the natural frequencies.'' This maximum likelihood (ML) estimation strategy is probabilistic in nature and requires an a priori model of the statistical distribution of the masses and stiffnesses of the blades. Proceeding with a preliminary assessment of the reliability of this approach, a simple Gaussian distribution of these parameters was assumed and the identification problem was found to be equivalent to the minimization of a simple, quadratic cost function, Eq. (7). Additionally, three types of constraints were introduced, the first one of which is required and is associated with the matching of the observed natural frequencies by the identified model (frequency constraints). The remaining two sets of constraints are optional and were described as modeling and mode shapes constraints, respectively. The former conditions were shown to be associated with fixed relationships between the structural parameters of the blade that are associated with its assumed dynamic model. On the contrary, the latter type of constraints is characterized by coefficients that vary from blade to blade and is encountered in particular when requiring that the blade model exhibits a measured or specified mode shape. Although the frequency constraints are intrinsically nonlinear, it was shown that the determination of the structural parameters according to the ML method with these constraints can generally be accomplished noniteratively by solving one linear system of equations, Eq. (15), although a few iterations are recommended when the natural frequencies of the blades alone are close together.

A preliminary parametric study of the reliability of the ML method has shown that this procedure yields generally very reliable estimates of the forced response of mistuned bladed disks, in contrast to the RMS approach. A more detailed assessment of the ML method is presented in the companion paper (Part II).

The availability of mode shape constraints enabled the assessment of the importance of mode shape mistuning on the forced response of bladed disks, a seldom discussed issue. In fact, comparisons were made between the reliability of the forced response predictions corresponding to identified blade models in which only the modal stiffnesses vary from blade to blade (the RMS approach), both modal stiffnesses and masses fluctuate (the ML approach with mode constraints), and modes shapes as well as modal parameters are mistuned (unconstrained ML approach). The results of this analysis indicate that the inclusion and accurate estimation of the blade-to-blade variability of their mode shapes are essential to accurately predict the forced response of mistuned disks from measurements of the corresponding blade alone natural frequencies.

Acknowledgments

The financial support of this investigation by a contract from the General Electric Aircraft Engine Company is gratefully acknowledged. The authors also wish to thank Dr. R. E. Kielb for insightful discussions of this problem.

References

[1] Wei, S. T., and Pierre, C., 1988, ''Localization Phenomena in Mistuned Assemblies with Cyclic Symmetry. Part I. Free Vibrations,'' ASME J. Vib., Acoust., Stress, Reliab. Des., **110**, pp. 429–438.
[2] Griffin, J. H., and Hoosac, T. M., 1984, ''Model Development and Statistical Investigation of Turbine Blade Mistuning,'' ASME J. Vib., Acoust., Stress, Reliab. Des., **106**, pp. 204–210.
[3] Kielb, R. E., and Kaza, K. R. V., 1984, ''Effects of Structural Coupling on Mistuned Cascade Flutter and Response,'' ASME J. Eng. Gas Turbines Power, **106**, pp. 17–24.
[4] Basu, P., and Griffin, J. H., 1986, ''The Effect of Limiting Aerodynamic and Structural Coupling in Models of Mistuned Bladed Disk Vibration,'' ASME J. Vib., Acoust., Stress, Reliab. Des., **108**, pp. 132–139.
[5] Lin, C. C., and Mignolet, M. P., 1996, ''Effects of Damping and Damping Mistuning on the Forced Vibration Response of Bladed Disks,'' J. Sound Vib., **193**, pp. 525–543.
[6] Sinha, A., 1986, ''Calculating the Statistics of Forced Response of a Mistuned Bladed Disk Assembly,'' AIAA J., **24**, pp. 1797–1801.
[7] Wei, S. T., and Pierre, C., 1988, ''Localization Phenomena in Mistuned Assemblies with Cyclic Symmetry. Part II. Forced Vibrations,'' ASME J. Vib., Acoust., Stress, Reliab. Des., **110**, pp. 439–449.
[8] Sinha, A., and Chen, S., 1989, ''A Higher Order Technique to Compute the Statistics of Forced Response of a Mistuned Bladed Disk,'' J. Sound Vib., **130**, pp. 207–221.
[9] Castanier, M. P., Ottarson, G., and Pierre, C., 1997, ''A Reduced Order Modeling Technique for Mistuned Bladed Disks,'' ASME J. Vibr. Acoust., **119**, pp. 439–447.
[10] Yang, M. T., and Griffin, J. H., 1997, ''A Reduced Order Approach for the Vibration of Mistuned Bladed Disk Assemblies,'' ASME J. Eng. Gas Turbines Power, **119**, pp. 161–167.
[11] Mignolet, M. P., Lin, C. C., and LaBorde, B. H., 2001, ''A Novel Limit Distribution for the Analysis of Randomly Mistuned Bladed Disks,'' ASME J. Eng. Gas Turb. Power, **123**, pp. 388–394.
[12] Lin, C. C., and Mignolet, M. P., 1997, ''An Adaptive Perturbation Scheme for the Analysis of Mistuned Bladed Disks,'' ASME J. Eng. Gas Turbines Power, **119**, pp. 153–160.
[13] Mignolet, M. P., and Hu, W., 1998, ''Direct Prediction of the Effects of Mistuning on the Forced Response of Bladed Disks,'' ASME J. Eng. Gas Turbines Power, **120**, pp. 626–634.
[14] Mignolet, M. P., and Lin, C. C., 1997, ''Identification of Structural Parameters in Mistuned Bladed Disks,'' ASME J. Vibr. Acoust., **119**, pp. 428–438.
[15] Benjamin, J. R., and Cornell, C. A., 1970, *Probability, Statistics, and Decision for Civil Engineers*, McGraw-Hill, New York.
[16] Lutes, L. D., and Sarkani, S., 1997, *Stochastic Analysis of Structural and Mechanical Vibrations*, Prentice–Hall, Englewood Cliffs.
[17] Rivas-Guerra, A. J., 1997, ''Identification of Mistuning Characteristics in Reduced Order Models of Bladed Disks,'' M.S. thesis, Arizona State University.
[18] Delor, J. P., 1998, ''Formulation and Assessment of Identification Strategies of Mistuned Bladed Disk Models from Free Response Data,'' M.S. thesis, Arizona State University.
[19] Thomas, D. L., 1979, ''Dynamics of Rotationally Periodic Structures,'' Int. J. Numer. Methods Eng., **14**, pp. 81–102.

Identification of Mistuning Characteristics of Bladed Disks From Free Response Data—Part II

A. J. Rivas-Guerra

M. P. Mignolet

Department of Mechanical and Aerospace Engineering, Arizona State University, Tempe, AZ 85287-6106

J. P. Delor

Vehicle Dynamics and Mechanisms, General Motors Corporation, 2000 Centerpoint Parkway, Pontiac, MI 48341-3147

The focus of the present two-part investigation is on the estimation of the dynamic properties, i.e., masses, stiffnesses, natural frequencies, mode shapes and their statistical distributions, of turbomachine blades to be used in the accurate prediction of the forced response of mistuned bladed disks. As input to this process, it is assumed that the lowest natural frequencies of the blades alone have been experimentally measured, for example in a broach block test. Since the number of measurements is always less than the number of unknowns, this problem is indeterminate in nature. In this second part of the investigation, the maximum likelihood method (ML) will first be revisited and a thorough assessment of its reliability in a wide variety of conditions, including the presence of measurement noise, different distributions of blade structural properties, etc., will be conducted. Then, a technique that provides a bridge between the two identification methods introduced in Part I, i.e., the random modal stiffnesses (RMS) and maximum likelihood (ML) approaches, will be presented. This technique, termed the improved random modal stiffnesses (IRMS) method is based on the maximum likelihood concepts but yields a mistuning model similar to that of the random modal stiffnesses technique. Finally, the accuracy of the RMS, ML, and IRMS methods in predicting the forced response statistics of mistuned bladed disks will be investigated in the presence of close blade alone natural frequencies. [DOI: 10.1115/1.1338950]

Detailed Assessment of the ML Approach

In regards to the maximum likelihood (ML) approach, the focus of the first part of this investigation was on both the formulation and the overall assessment of the methodology, see Figs. (I.3)–(I.7).[1] To provide a confirmation of the reliability of the ML approach, its behavior in situations that do not conform to the assumptions made during its formulation should be investigated. Specifically, it is necessary to evaluate the consequences of choosing a covariance matrix that does not closely match its exact counterpart, to investigate the reliability of the ML approach when the original population of stiffnesses and masses follows a non-Gaussian distribution, to assess the influence of measurement noise in the natural frequencies and the effects of unobserved blade alone natural frequencies.

Effects of Incorrect Assumed Covariance Matrix. It was noted in Part I that the maximum likelihood approach relies on the availability of the covariance matrix $K_{\mathbf{XX}}$, see Eqs. (I.6) and (I.15). Since the maximum likelihood principle is used to solve a problem which otherwise would be indeterminate, it should not be expected that this covariance matrix can be obtained solely from the estimation process.

As the matrix $K_{\mathbf{XX}}$ cannot be obtained from the measured frequencies, an estimate of this quantity must be selected from prior information about the process(es) that created the randomness of the stiffnesses and masses, i.e., the manufacturing and/or the in-service wear. An attractive way to obtain such a prior estimate is to quantify the fluctuations of the stiffnesses and masses that result from the randomness in the geometrical and materials properties of the blades. This task can in particular be achieved by relying on the techniques developed by Collins and Thomson [1], Shinozuka and Astill [2], and Vaicaitis [3].

Note further in relation to the determination of $K_{\mathbf{XX}}$ that this matrix can be multiplied by an arbitrary scalar constant without affecting the value of the vector $\mathbf{X}$ that minimizes Eq. (I.7). In this light, it is not necessary to estimate the exact numerical values of the elements of the covariance matrix $K_{\mathbf{XX}}$, only their relative values with respect to a particular element of this matrix are required.

Consider for example the one-degree-of-freedom per blade model shown in Fig. (I.2) in which the vector $\mathbf{X}$ was chosen as

$$\mathbf{X}=\begin{bmatrix} k_j \\ \omega_t^2\ m_j \end{bmatrix}, \tag{1}$$

where $\omega_t^2=k_t/m_t$ is the tuned blade alone natural frequency. Accordingly, the covariance matrix $K_{\mathbf{XX}}$ is

$$K_{\mathbf{XX}}=\begin{bmatrix} \sigma_k^2 & \rho_{km}\sigma_k\omega_t^2\sigma_m \\ \rho_{km}\sigma_k\omega_t^2\sigma_m & \omega_t^4\sigma_m^2 \end{bmatrix}=\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}10^{-4}k_t^2 \tag{2}$$

since the coefficient of correlation of the masses and stiffnesses, ρ_{km}, vanishes in view of their assumed statistical independence. Using the above form of the covariance matrix in Eq. (I.15) with one frequency constraint, i.e., $m=1$, led to an excellent forced response prediction for all engine orders, see Fig. I.3.

Then, note that the exact covariance matrix, Eq. (2), can be written as $K_{\mathbf{XX}_{\text{exact}}}=\left[\begin{smallmatrix}1 & 0\\ 0 & 2\end{smallmatrix}\right]$ by a multiplication by $1/\sigma_k^2$. This observation demonstrates that the structural estimates only depend on the ratio of variances $(\omega_t^4\sigma_m^2)/\sigma_k^2$, not on the variances of the masses and stiffnesses, σ_m^2 and σ_k^2, separately. Equivalently, the covariance matrix specifies only the **relative** variations and correlations between masses and stiffnesses.

[1]To avoid duplication, all equations and figures marked with a roman numeral I in this part will refer to the equations and figures with the same number in Part I.

Contributed by the Structures & Dynamics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the S&D Division, May 4, 1999; final revision received by the ASME Headquarters June 9, 1999. Editor: H. D. Nelson.

Table 1 Errors in the mean and in the standard deviation of the amplitude of the highest responding blade for different assumed covariance matrices (ML approach) and for the RMS approach, single-degree-of-freedom blade model excited at the tuned natural frequency corresponding to the ninth engine order, ω_9^t.

	ML, Ex. K_{XX}, $\omega_t^4\sigma_m^2/\sigma_k^2=2$	ML, Inc. K_{XX}, $\omega_t^4\sigma_m^2/\sigma_k^2=1$	ML, Inc. K_{XX}, $\omega_t^4\sigma_m^2/\sigma_k^2=4$	RMS
Mean	−0.15%	−0.73%	0.29%	−2.65%
Stand. Dev.	−0.35%	−4.11%	2.61%	−15.78%

Table 2 Errors in the mean and in the standard deviation of the amplitude of the highest responding blade, mass m_2, for different assumed covariance matrices (ML approach) and for the RMS approach, two-degree-of-freedom blade model 1 excited at the lowest tuned natural frequency corresponding to the ninth engine order, $\omega_{9,1}^t$, both blade alone natural frequencies measured.

	ML, Ex. K_{XX}, Eq. (4)	ML, Inc. K_{XX}, K_{XX}=Ident.	RMS
Mean	−0.023%	0.188%	−1.092%
Stand. Dev.	0.024%	1.389%	−7.674%

To test the robustness of the ML approach, two different incorrect covariance matrices were used and the corresponding forced response statistics were computed. In the first case, the ratio of the variances was selected to be half of its exact value, i.e., $\omega_t^4\sigma_m^2/\sigma_k^2=0.5(\omega_t^4\sigma_m^2/\sigma_k^2)_{\text{exact}}$, resulting in a covariance matrix $K_{XX}=\left[\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right]$. In the other one, on the contrary, this ratio was chosen to be twice as large, i.e., $\omega_t^4\sigma_m^2/\sigma_k^2=2(\omega_t^4\sigma_m^2/\sigma_k^2)_{\text{exact}}$, resulting in a covariance matrix $K_{XX}=\left[\begin{smallmatrix}1 & 0\\ 0 & 4\end{smallmatrix}\right]$. Shown in Table 1 is the behavior of the errors in the statistics of the forced response as functions of $\omega_t^4\sigma_m^2/\sigma_k^2$. These results indicate, as expected, that the errors in mean and standard deviation are larger when an incorrect covariance matrix is selected than they are when the exact covariance matrix is used ($\omega_t^4\sigma_m^2/\sigma_k^2=2$). However, the errors obtained with these incorrect covariance matrices are still much smaller than the ones corresponding to the RMS approach, see Table 1. These results demonstrate that the forced response of the bladed disk can be very well estimated even if the assumed covariance matrix is not exact.

In assessing the effects of the assumed covariance matrix, it should further be noted that the RMS technique is in fact a particular case of the ML approach in which the assumed covariance matrix has been selected as $K_{XX}=\left[\begin{smallmatrix}1 & 0\\ 0 & 0\end{smallmatrix}\right]$. This observation suggests that the RMS approach can be viewed as a limiting or ''worse case'' version of the ML strategy.

As a final check of the mildness of the effects of selecting an incorrect covariance matrix, consider the two-degree-of-freedom per blade model (model 1) shown in Fig. I.1. Defining the vector **X** as

$$\mathbf{X}=[k_1 \;\; k_2 \;\; \omega_d^2 m_1 \;\; \omega_d^2 m_2]^T, \tag{3}$$

where ω_d^2 is the average of the squared natural frequencies of the tuned blades, leads to the exact covariance matrix

$$K_{XX}=\begin{bmatrix} 88.82 & 0.30 & 0 & 0 \\ 0.30 & 0.30 & 0 & 0 \\ 0 & 0 & 46.85 & 0 \\ 0 & 0 & 0 & 1.00 \end{bmatrix}, \tag{4}$$

after scaling by the original element (4,4). Then, to assess the effect of an incorrect covariance matrix, it was assumed that K_{XX} is the identity matrix, which is clearly very different from the exact one, Eq. (4). The results of the corresponding identification, see Table 2, confirm the above-mentioned findings that the errors in the predicted response of the bladed disk are affected by selecting an incorrect covariance matrix. However, even with a broadly incorrect covariance matrix, the ML strategy stills performs much better than the RMS approach. Note finally that the existence of a minimum of the errors in the predicted statistics of the forced response when the assumed covariance matrix K_{XX} is close to its exact counterpart could be used, in connection with experimental data, to estimate the exact K_{XX}.

Effects of a Non-Gaussian Distribution of the Structural Parameters. In the formulation of the ML strategy, it was assumed that the masses and stiffnesses follow a jointly Gaussian distribution, Eq. (I.6). An assumption on the distribution of these parameters was required to obtain an expression for the probability density function of the vector **X**. Further, the specific choice of the Gaussian distribution was motivated by the frequent appearance of this distribution in practical problems (resulting from the central limit theorem) but also by the simple mathematical optimization problem that leads to Eq. (I.7). The excellent results obtained in the previous sections in connection with masses and stiffnesses that follow a Gaussian distribution justify the assessment of the ML approach when the population of masses and stiffnesses is not normally distributed. To achieve this task, the blades properties of the single-degree-of freedom model (see Fig. I.2) were first simulated as uniformly distributed random variables, then the natural frequencies corresponding to these populations of blade properties were computed and used as input to the ML approach (still based on the Gaussian assumption, Eq. (I.6)) to estimate the blade stiffnesses and masses. Once these parameters were obtained, the forced response of a series of mistuned bladed disk was determined. Shown in Fig. 1 are the corresponding probability density functions of the amplitude of the highest responding blade for a ninth engine order excitation, $r=9$. Note again that the ML curve matches very well its exact counterpart, much better than the random modal stiffnesses (RMS) prediction does. This comparison

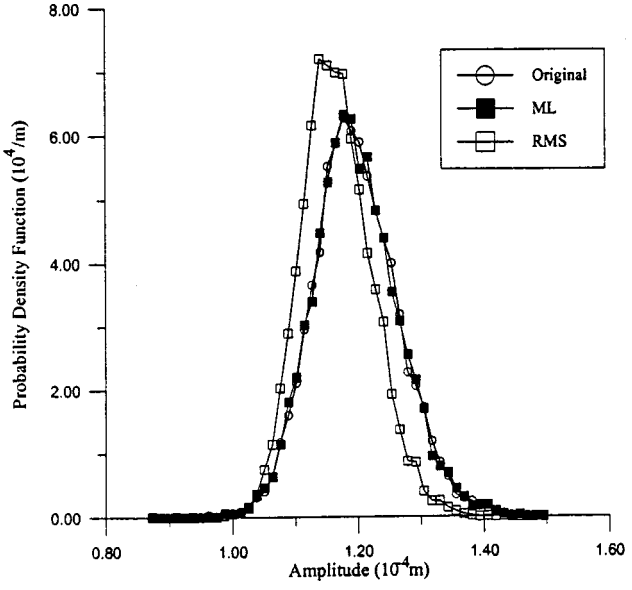


Fig. 1 Probability density function of the amplitude of the highest responding blade. Original parameters, estimated by ML and RMS approaches, single-degree-of-freedom blade model excited at the tuned natural frequency corresponding to the ninth engine order ω_9^t, uniform distribution of blade parameters.

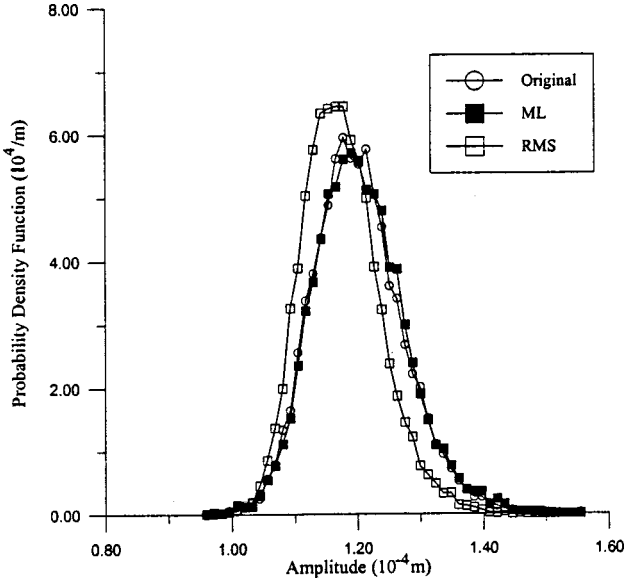


Fig. 2 Probability density function of the amplitude of the highest responding blade. Original parameters, estimated by ML and RMS approaches, single-degree-of-freedom blade model excited at the tuned natural frequency corresponding to the ninth engine order ω_9^t, measurement noise of standard deviation=0.25 percent.

demonstrates that the ML can reliably be used to predict the forced response of mistuned bladed disks regardless of the distribution of the mass and stiffness mistuning.

Effects of Measurement Noise (Errors in the Measurement of Frequencies). No measurement system has a perfect accuracy. When physical variables, the natural frequencies of the blades in the present context, are measured there always exists a certain margin of error which can adversely affect the reliability of the data analysis/estimation process results. In this light, it is necessary to investigate the influence of small errors in the measurement of the blade natural frequencies on the prediction of the forced response of the bladed disks. In order to accomplish this goal an artificial noise factor was introduced in the computation of the simulated ''experimental'' natural frequencies. That is, for the one-degree-of-freedom per blade model these values were first obtained as

$$\omega_j = \sqrt{\frac{k_j}{m_j}}(1+\epsilon_\omega), \tag{5}$$

where ϵ_ω, which represents the measurement noise, is a normal random variable with zero mean and standard deviation σ_ϵ. Then, both the ML and the RMS approaches were used to estimate the structural parameters from these noisy natural frequencies from which in turn the forced response of the entire disk was computed. As demonstrated by the probability density functions presented in Fig. 2, the reliability of the ML and RMS approaches in predicting the forced response does not substantially change due to the presence of a measurement noise of standard deviation $\sigma_\epsilon = 0.25$ percent. The corresponding results for a standard deviation of 0.50 percent, not shown here for brevity, demonstrate that large errors in the prediction of the forced response can occur. Shown in Table 3 are the errors in the mean and standard deviation of the forced response at a zeroth engine order excitation for different levels of measurement noise. On the basis of the above results, it is concluded that the measurement of the natural frequencies should be accurate to within 0.50 percent, preferably 0.25 percent, to avoid noise corruption. This recommendation corresponds to the particular case where $\sigma_k = 0.01k_t$, $\sigma_m = 0.01\sqrt{2}m_t$ and $\rho_{km} = 0$ (see Fig. I.2) for which the standard deviation of the natural frequencies due to mistuning only is $\sigma_\omega = 0.01(\sqrt{3}/2)\omega_t$. In this light, a standard deviation of the noise, σ_ϵ, of 0.1 percent corresponds to a signal-to-noise ratio (SNR) of $5\sqrt{3}=8.66$ when the signal is considered to be the fluctuations in frequency associated with mistuning. Similarly, standard deviations of the noise of 0.25 and 0.50 percent are equivalent to SNR=3.46 and 1.73. Then, the results shown in Table 3 indicate that the ML technique yields reliable estimates of the forced response statistics provided that the measurement errors on the natural frequencies do not exceed 1/4 to 1/3 of the variability of these frequencies due to mistuning.

Table 3 Errors in the mean and in the standard deviation of the amplitude of the highest responding blade for different measurement noise strength, ML approach, single-degree-of-freedom blade model excited at the tuned natural frequency corresponding to the zeroth engine order, ω_0^t.

std(noise)	0.00 % (SNR=∞)	0.10% (SNR=8.66)	0.25 % (SNR=3.46)	0.50 % (SNR=1.73)
Mean	-0.003 %	0.067 %	0.397 %	1.346 %
Stand. Dev.	-0.012 %	0.454 %	3.445 %	11.215 %

Table 4 Errors in the mean and in the standard deviation of the amplitude of the highest responding blade, mass m_2, when either one or the two blade alone natural frequencies are measured, two-degree-of-freedom blade model 1 excited at the lowest tuned natural frequency corresponding to the ninth engine order, $\omega_{9,1}^t$.

	ML One Freq.	RMS One Freq.	ML Two Freq.	RMS Two Freq.
Mean	-0.052 %	-1.217 %	-0.023 %	-1.092 %
Stand. Dev.	-0.290 %	-8.650 %	0.024 %	-7.674 %

Effects of Unobserved Blade Frequencies. Finite element methods, which are widely used in practice, yield blade models that possess hundreds of degrees of freedom and are thus impractical for mistuning analyses. Fortunately, reduced order modeling strategies [4,5] have been proposed that provide a dynamic response prediction of accuracy equivalent to that of a full finite element model but with a much smaller number of degrees of freedom, 10 to 20 say. It is, however, not expected that the 10 to 20 natural frequencies needed to identify such models will be reliably determined from a broach block test so that a lack of measured natural frequencies is also expected to occur. In the RMS estimation procedure, this further indetermination is resolved by enforcing additional constraints, i.e., the unknown mistuned frequencies are replaced by the tuned frequencies of those particular modes. The ML technique is, however, designed to work with indetermination. Thus, the only difference in the ML procedure lies in the fewer number of constraints associated with the minimization of Eq. (I.7). Although the full dynamic model of the blades can still be estimated, note that no information exists on the variability of the modal impedance of the modes whose frequencies were not measured. Thus, it should be expected that the forced response of mistuned bladed disks at these frequencies will not be predicted properly.

To assess the reliability of both the RMS and ML estimation strategies in the presence of unobserved frequencies, the structural parameters of the two-degree-of-freedom blade model (model 1) were determined assuming that the second frequency of each blade, ω_2, was not recorded. Once the structural parameters of the blades were computed, the disk was excited in both its low and high frequency bands and the statistics of the corresponding forced response were compared to their exact values. As expected the results of neither the ML nor the RMS strategy were satisfactory in the high frequency band (not shown here). On the contrary, the behavior of both estimation methods in the low, observed, frequency band was found to be very similar to the situation in which the two frequencies were known. A summary of the results,

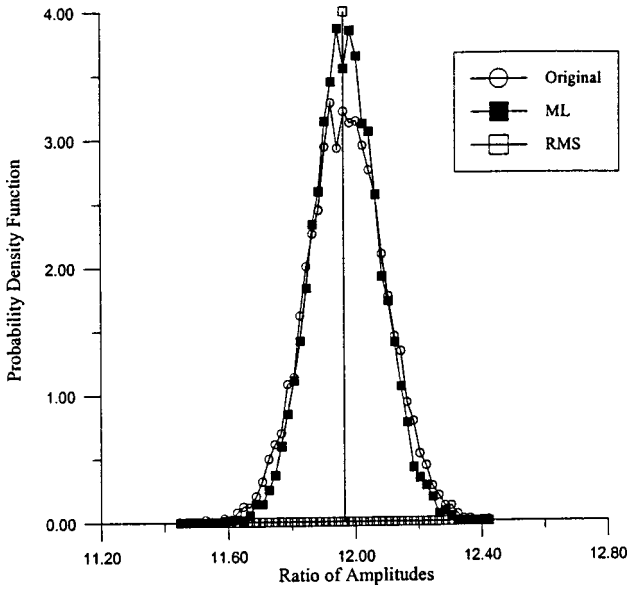


Fig. 3 Probability density function of the ratio of the modal displacements of masses 1 and 2, low frequency mode. Original parameters, estimated by ML and RMS approaches, two-degree-of-freedom blade model 1.

presented in Table 4, indicates only a slight degradation in the accuracy of the predicted statistics of the forced response.

Estimation of the Variations in the Mode Shapes. The importance of estimating the blade alone mode shape variability was demonstrated in Part I by comparing the ML results obtained with and without mode shape constraints, see Fig. (I.5). The excellent reliability of the unconstrained ML results shown would indicate that this approach has well captured the blade alone mode shape variability. To confirm this expectation, the probability density function of the modal response ratio mass 2/mass 1 was determined for both the original population of blade stiffnesses and masses and the one identified by the ML approach (without mode shape constraints). The results of this investigation are plotted in Fig. 3 for the mode shape associated with the lowest natural frequency; a similar accuracy was also observed for the high frequency mode. The good matching between the two curves shown in Fig. 3 indicates that the ML approach can also provide accurate estimates of the variability of the mode shapes of the blades.

This finding is very positive but it is also quite surprising and it is necessary to clarify how and how well mode shape variability can be extracted from the sole knowledge of the fundamental natural frequencies of the blades. To this end, note first that the variation of any structural parameter will in general produce a change in both the natural frequencies and the corresponding mode shapes. Equivalently, the variations in frequencies and modes are linked to each other through the fluctuations of the structural parameters that created them. If this link is ''strong'' enough or, in a statistical terminology, if the coefficient of correlation between the random changes in natural frequencies and modes is large enough, the knowledge of the former will provide a good basis to estimate the latter.

An approximate expression for the coefficient of correlation between natural frequencies and mode shapes can be obtained by proceeding with perturbation techniques. Specifically, denote first by K_t, M_t, ω_j^t, and $\boldsymbol{\phi}_j^t$ the tuned stiffness and mass matrices and their corresponding natural frequencies and mass normalized mode shapes, $j=1,2,\ldots,N$. Clearly, these quantities satisfy the eigenvalue problem

$$K_t\boldsymbol{\phi}_j^t=(\omega_j^t)^2M_t\boldsymbol{\phi}_j^t \quad \text{and} \quad (\boldsymbol{\phi}_j^t)^TM_t\boldsymbol{\phi}_l^t=\delta_{jl}, \tag{6}$$

where δ_{jl} denotes the Kronecker symbol. Next, let the variations of these properties due to mistuning be

$$\delta K=K-K_t; \quad \delta M=M-M_t; \quad \delta\omega_j^2=\omega_j^2-(\omega_j^t)^2 \quad \text{and} \quad \delta\boldsymbol{\phi}_j=\boldsymbol{\phi}_j-\boldsymbol{\phi}_j^t \tag{7}$$

and note that they satisfy the perturbed eigenvalue problem

$$(K_t+\delta K)(\boldsymbol{\phi}_j^t+\delta\boldsymbol{\phi})=[(\omega_j^t)^2+\delta\omega_j^2](M_t+\delta M)(\boldsymbol{\phi}_j^t+\delta\boldsymbol{\phi}) \tag{8a}$$

with the normalization condition

$$(\boldsymbol{\phi}_j^t+\delta\boldsymbol{\phi})^TM_t(\boldsymbol{\phi}_j^t+\delta\boldsymbol{\phi})=1. \tag{8b}$$

Then, assuming a representation of the mode shape perturbation in the form

$$\delta\boldsymbol{\phi}_j=\sum_{i=1}^{N}\alpha_{ji}\boldsymbol{\phi}_i^t \tag{9}$$

and keeping first-order terms only in Eq. (8), lead to the expressions $\alpha_{jj}=0$ and

$$\alpha_{ji}[(\omega_i^t)^2-(\omega_j^t)^2]=(\boldsymbol{\phi}_i^t)^T[(\omega_j^t)^2\delta M-\delta K](\boldsymbol{\phi}_j^t) \quad i\neq j \tag{10}$$

and

$$\delta\omega_j^2=-(\boldsymbol{\phi}_j^t)^T[(\omega_j^t)^2\delta M-\delta K](\boldsymbol{\phi}_j^t). \tag{11}$$

Using these relations, the statistical moments of the coefficients α_{ji} and of the variations of the natural frequencies $\delta\omega_j^2$ can be determined in terms of the corresponding moments of the stiffness and mass mistuning matrices δK and δM, if they are known. In particular, in the likely case of zero mean fluctuations of the structural parameters, it is found that the coefficient of correlation between the change in natural frequency $\delta\omega_j^2$ and the mode shape variation expansion coefficient α_{ji} can be expressed as

$$\text{corr}(\alpha_{ji},\delta\omega_j^2)=\frac{E[\alpha_{ji}\delta\omega_j^2]}{\sqrt{E[\alpha_{ji}^2]E[(\delta\omega_j^2)^2]}}. \tag{12}$$

Thus, following the above discussion, reliable estimates of the mode shape variability and consequently of the subsequent forced response prediction should be obtained when $\text{corr}(\alpha_{ji},\delta\omega_j^2)$ is large. This argument is corroborated by the analysis of the two-degree-of-freedom per blade model shown in Fig. I.1 (model 1) which yielded excellent mode shape estimates (see Fig. 3) and is associated with $\text{corr}(\alpha_{12},\delta\omega_1^2)=0.551$.

To further validate the adequacy of the above coefficient of correlation as an indicator of the reliability of the identification process, a second two-degree-of-freedom per blade model, defined as model 2 in Fig. I.1, was constructed that is of the same form as model 1 but with different tuned values. For this rather arbitrary model, the coefficient of correlation was found to be $\text{corr}(\alpha_{12},\delta\omega_1^2)=0.116$. The small value of this quantity suggests that the statistical distribution of the mode shapes obtained by using the ML method should not closely match its true counterpart. This expectation is confirmed in Fig. 4. Relying on the arguments developed in Part I, the worsening in mode shape prediction when switching from blade model 1 to 2 should be accompanied by a similar degradation in the reliability of the forced response estimates. Shown in Fig. 5 are the errors in the mean and standard deviation of the response of mass m_2 as functions of the engine order and corresponding to both the ML and RMS methods. Note indeed that both ML errors are larger than those shown in Fig. I.5 but also that they are still smaller than their RMS counterparts which incidentally always correspond to $\text{corr}(\alpha_{12},\delta\omega_1^2)=0$. The above comments suggest that the correlation between α_{ji} and $\delta\omega_j^2$, Eq. (12), can be used as a measure of the reliability of the ML method in predicting the forced response of mistuned bladed disks.

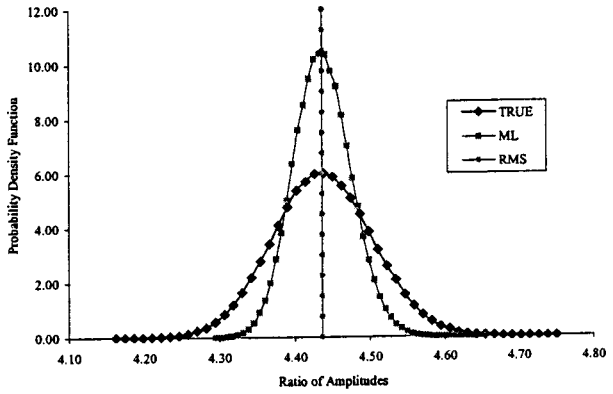


Fig. 4 Probability density function of the ratio of the modal displacements of masses 1 and 2, low frequency mode. Original parameters, estimated by ML and RMS approaches, two-degree-of-freedom blade model 2.

In practical applications, it is unclear that the statistical moments of the stiffness and mass variations would be known a priori, see the earlier comments regarding the covariance matrix $K_{\mathbf{XX}}$. However, the values of these moments obtained by application of the ML method can be used as a basis to estimate the coefficient of correlation and thus to provide a measure of confidence on the forced response prediction. In fact, performing these computations on the two models led to $\mathrm{corr}(\alpha_{12}, \delta\omega_1^2) = 0.640$ for model 1 and 0.206 for model 2, both of which agree well with the exact values of 0.551 and 0.116.

Randomized Maximum Likelihood. It was observed in the first part of this investigation that the ML method leads, by formulation, to an underprediction of the level of mistuning in the blade structural properties. In turn, this underestimation of the variability of the blade masses and stiffnesses provokes a similar error in the statistics of the amplitude of forced response and a shift to the left of the corresponding probability density function, both of which were noted in Part I. These findings suggest that the introduction in the ML formulation of a bias toward higher mistuning levels might be beneficial. Clearly, however, this bias cannot be deterministic as that would correspond to a shift in the tuned system.

The procedure suggested here involves first the generation of a random target vector $\mathbf{X}_T$ of mass and stiffness deviations for each blade. Then, the randomized maximum likelihood estimates of the blade properties stacked in the vector $\mathbf{X}$ minimize the error

$$\varepsilon_{\mathrm{RML}} = (\mathbf{X} - \mathbf{X}_T)^T K_{\mathbf{XX}}^{-1} (\mathbf{X} - \mathbf{X}_T) \tag{13}$$

while satisfying frequency, modeling, and mode shape constraints given by Eqs. (I.5), (I.8), and (I.9).

To test the reliability of this randomized ML strategy, the one-degree-of-freedom per blade model of Fig. I.2 was reconsidered with $\sigma_k/k_t = \sigma_m/m_t = 0.01$. The random components k_T and m_T of the target vector $\mathbf{X}_T = [k_T \ \omega_t^2 m_T]^T$ were selected as independent random variables with zero means and respective standard deviations σk_t and σm_t. Then, shown in Fig. 6 are the errors in the mean and standard deviation of the amplitude of the highest responding blade as function of σ for a 12th engine order excitation at the corresponding tuned natural frequency. For $\sigma = 0$, the regular ML method is recovered and small negative errors in the forced response prediction are obtained. More importantly however, note that both errors practically vanish for $\sigma = 0.01$ and an extremely accurate prediction of the forced response is obtained. These results and others, not shown here for brevity, demonstrate that very low errors in the forced response statistics are typically obtained when the standard deviations of the target properties, i.e., of the vector $\mathbf{X}_T$, are close to their exact counterparts. In this light, it is recommended that the covariance matrix of this random vector be selected to match the one used in the ML formulation, i.e., $K_{\mathbf{XX}}$.

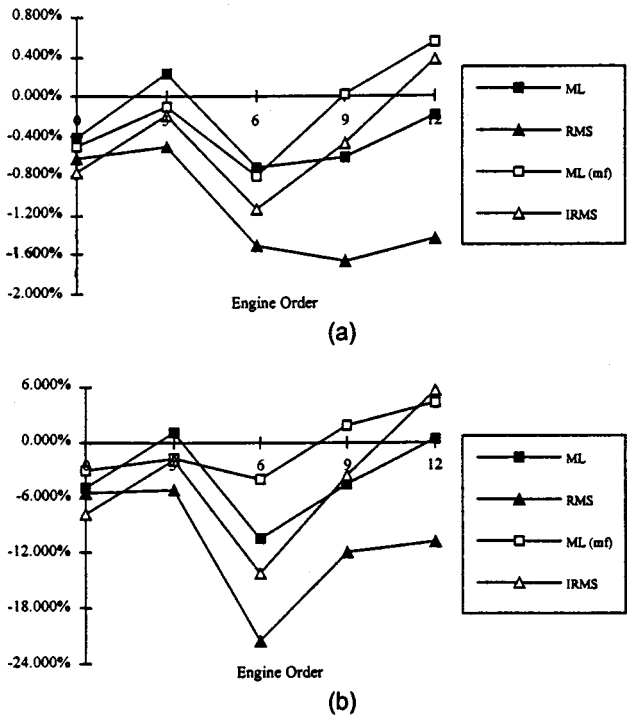


Fig. 5 Errors (*a*) in the mean and (*b*) in the standard deviation of the maximum response of the mass m_2. ML, ML with first mode fixed (mf), RMS and IRMS strategies, two-degree-of-freedom blade model 2 excited at the lowest tuned natural frequency corresponding to the *r*th engine order, $\omega_{r,1}^t$.

Improved Random Modal Stiffness (IRMS) Method

Introduction. It was demonstrated in the first part of this investigation that the RMS approach provides a model of the blades that is appropriate for blade alone analyses but does not correctly represent the dynamic behavior of the blade as part of the disk. Since the prediction of the response of the entire disk is ultimately the purpose of the present identification effort, it can be questioned whether this approach can be modified to focus on the blade as a substructure of the disk system as opposed to a structure on its own.

To accomplish this task, it is suggested here to *select the modal stiffnesses of the blades, keeping their modal masses and mode shapes equal to their tuned values, so that the natural frequencies of the corresponding uniformly mistuned disk in the selected engine order, not of the blades alone, be exactly matched.*

An inspection of Eq. (I.23) immediately reveals a difficulty in this effort: Since the blade structural parameters are not known, the natural frequencies of the corresponding uniformly mistuned disk cannot be predicted in advance! To palliate this situation, it is proposed here to proceed in two steps. First, the ML approach is used to produce estimates of the masses and stiffnesses of the blades from which the unknown natural frequencies of uniformly mistuned disks can themselves be estimated. Second, the modal stiffnesses of the actual blade model are then selected to exactly match these system-wide frequencies.

In this light, this new RMS-type formulation cannot be considered as an independent identification approach since it uses the results of the ML approach. Rather, it may be considered as a ''model simplification'' technique in which the randomness is concentrated on the modal stiffnesses of the blades. The benefits of investigating such a strategy are twofold. First, the analysis will

provide additional information on the appropriateness of a modal-stiffness-only modeling of mistuning as well as on the validity of Eq. (I.30) to clarify the issues associated with this modeling. Second, this technique could be used to adapt the ML methodology to situations that privilege the RMS approach.

Formulation. According to the above strategy, the general formulation of the proposed technique, termed the improved random modal stiffnesses (IRMS) method, proceeds as follows. First, estimates of the blade stiffness and mass matrices K and M are obtained by the ML method. Then, a tuned disk is considered in which each blade is characterized by the identified stiffness and mass matrices K and M, respectively. This system possesses N natural frequencies that are associated with the rth engine order, i.e., $\omega_{r,j}$, where $j=1,2,\ldots,N$. From Eq. (I.23), these frequencies are solutions of the characteristic equation

$$\det[(K+K^{(r)})-\omega_{r,j}^2(M+M^{(r)})]=0 \quad j=1,\ldots,N; \quad \text{and} \quad r \text{ fixed.} \tag{14}$$

To meet the requirements and format of the IRMS method, the corresponding identified stiffness and mass matrices K^{IRMS} and M^{IRMS}, respectively, should satisfy the following conditions.

(1) Exhibit mode shapes and modal masses equal to their tuned counterparts. As in Eqs. (I.2)–(I.4), this condition is equivalent to the relations

$$M^{\mathrm{IRMS}}=\Phi_t^{-T}\Phi_t^{-1} \quad \text{and} \quad K^{\mathrm{IRMS}}=\Phi_t^{-T}\Lambda^{\mathrm{IRMS}}\Phi_t^{-1} \tag{15}$$

where Λ^{IRMS} denotes the diagonal modal stiffness matrix for the mistuned blade, i.e.,

$$\Lambda^{\mathrm{IRMS}}=\mathrm{diag}(\Lambda_{11},\Lambda_{22},\ldots,\Lambda_{NN}). \tag{16}$$

(2) Yield system natural frequencies corresponding to the engine order r that exactly match those predicted on the basis of the ML estimates of the stiffness and mass matrices. This last condition implies that the natural frequencies $\omega_{r,j}, j=1,2,\ldots,N$, determined from Eq. (14) must also satisfy the relation

$$\det[(K^{\mathrm{IRMS}}+K^{(r)})-\omega_{r,j}^2(M^{\mathrm{IRMS}}+M^{(r)})]=0. \tag{17}$$

Introducing Eq. (15) into Eq. (17) leads to the conditions

$$\det[\Lambda^{\mathrm{IRMS}}+A_j^{(r)}]=0 \quad j=1,2,\ldots,N; \quad r \text{ fixed}, \tag{18}$$

where the matrix $A_j^{(r)}$ of elements $A_{j,pq}^{(r)}$ is defined as

$$A_j^{(r)}=\Phi_t^T[K^{(r)}-\omega_{r,j}^2M^{(r)}]\Phi_t-\omega_{r,j}^2I_N. \tag{19}$$

Once the N equations (18) have been solved for the N elements Λ_{jj} of the diagonal matrix Λ^{IRMS}, the IRMS estimates of the mass and stiffness matrices of the blade can be computed by Eq. (15). In connection with a one-degree-of-freedom per blade model, these computations are readily accomplished and yield $m^{\mathrm{IRMS}}=m_t$ and

$$k^{\mathrm{IRMS}}=m_t\Lambda_{11}=m_t\omega_r^2-[k^{(r)}-\omega_r^2m^{(r)}]. \tag{20}$$

Solving Eq. (14) for the natural frequencies ω_r yields $\omega_r^2=(k+k^{(r)})/(m+m^{(r)})$ so that

$$k^{\mathrm{IRMS}}=\frac{(m_t+m^{(r)})k+(m_t-m)k^{(r)}}{(m+m^{(r)})}. \tag{21}$$

For the system shown in Fig. I.2, the above expression simplifies further to

$$k^{\mathrm{IRMS}}=\frac{m_t}{m}k+\frac{(m_t-m)}{m}\left[4k_C\sin^2\left(\frac{\pi r}{N_b}\right)\right]. \tag{22}$$

In the case of a two-degree-of-freedom system, Eq. (18) still admits a closed form solution as [6]

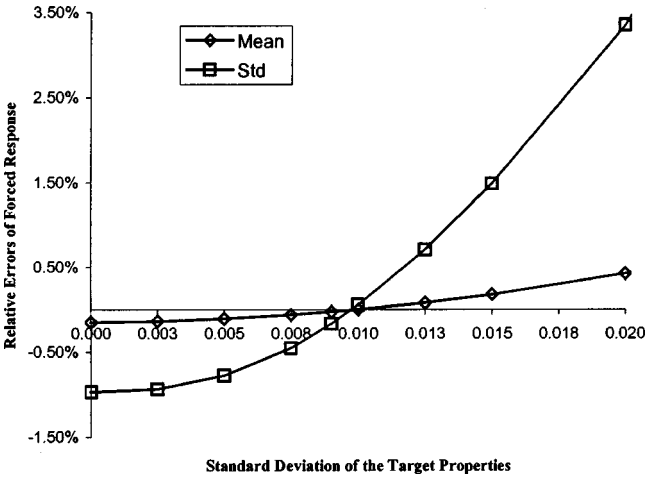


Fig. 6 Errors in the mean and in the standard deviation of the maximum amplitude of resonant response of the blade as functions of the standard deviation of the target properties. Randomized ML strategy, single-degree-of-freedom blade model excited at the tuned natural frequency corresponding to the 12th engine order, ω_{12}^t.

$$\Lambda_{11}=-\frac{1}{2}(A_{2,11}^{(r)}+A_{1,11}^{(r)}) \pm\frac{1}{2}\sqrt{(A_{2,11}^{(r)}-A_{1,11}^{(r)})^2-4\left(\frac{A_{2,11}^{(r)}-A_{1,11}^{(r)}}{A_{2,22}^{(r)}-A_{1,22}^{(r)}}\right)(A_{1,12}^{(r)})^2} \tag{23}$$

and

$$\Lambda_{22}=\frac{A_{2,11}^{(r)}A_{2,22}^{(r)}-A_{1,11}^{(r)}A_{1,22}^{(r)}+(A_{2,22}^{(r)}-A_{1,22}^{(r)})\Lambda_{11}}{A_{1,11}^{(r)}-A_{2,11}^{(r)}} \tag{24}$$

from which the masses and stiffnesses of the blade can be evaluated. It is important to note from Eq. (23) that the existence of the parameter Λ_{11} is not guaranteed since the term inside the square root is not always positive. Even when it is, the corresponding solutions Λ_{11} and Λ_{22} may not be greater than zero, which is required to obtain a positive definite stiffness matrix K^{IRMS}. Note also that when a bonafide solution does exist, it may not be unique because of the $\pm$ sign in Eq. (23). In most cases, however, the ''correct'' solution exists and is quite evident; it corresponds to values $\Lambda_{jj}\approx\omega_j^2$ as can be expected by comparing Eqs. (I.3) and (15). In the general case $N>2$, a closed form solution for the parameters Λ_{jj} was not found but an iterative numerical technique based on Newton's method was successfully employed (see Delor [6] for details).

It should finally be noted that the IRMS technique described above reduces exactly to the RMS approach when $r=0$ since the corresponding matrices $K^{(0)}$ and $M^{(0)}$ vanish identically. This result is consistent with the motivation behind the IRMS developments since a zeroth engine order excitation favors the blade alone behavior on which the RMS method is based.

Numerical Results. The reliability of the IRMS strategy was first assessed by considering the one-degree-of-freedom per blade model shown in Fig. I.2. Shown in Fig. I.3 are the relative errors in the mean and standard deviation of the corresponding maximum amplitude of resonant response on the disk as functions of the engine order r for the ML, RMS, and IRMS approaches. Clearly, the IRMS method leads to excellent estimates of the forced response statistics, its errors are typically much smaller than the ones corresponding to the RMS approach (except of course for $r=0$ where they are equal) and of the same order of magnitude as their ML counterparts.

Table 5 Blade parameters and modal characteristics for the four cases considered in the veering analysis Rel. Freq. Diff. $=2(\omega_2-\omega_1)/(\omega_2+\omega_1)$

Case	k_2 (N/m)	m_2 (kg)	Rel. Freq. Diff.	$corr(\alpha_{12}, \delta\omega_1^2)$
1	275.67	$7.313\ 10^{-6}$	2.54 %	0.0121
2	1,132.29	$3.010\ 10^{-5}$	5.14 %	0.0249
3	4,789.78	$1.284\ 10^{-4}$	10.58 %	0.0496
4	36,411.44	$1.055\ 10^{-3}$	29.74 %	0.1165

The results of a similar study conducted on the basis of the two-degree-of-freedom per blade models shown in Fig. I.1 (models 1 and 2) can be seen in Figs. I.5 and 5, respectively. Clearly, in both of these cases, the errors associated with the proposed IRMS method are typically close to their ML counterparts, except for $r=0$, and smaller than those corresponding to the RMS approach. Interestingly, it can be observed that the IRMS technique sometimes leads to an overprediction of the mean and standard deviation of the response which is in contradiction with the ML and RMS methods that consistently underpredict the values of these moments.

Identification of Mistuning in the Presence of Veering

Small changes in the values of the masses and stiffnesses of a dynamic system usually imply correspondingly small changes in its mode shapes. However, when two or more of its natural frequencies are close together, small perturbations of the system parameters can often lead to large variations of the corresponding mode shapes. This situation, referred to as mode veering, is in particular encountered in connection with weakly coupled bladed disks where a small mistuning can produce a dramatic localization of the mode shapes [7]. Note, however, that the veering phenomenon is not limited to the disk modes, it can also be encountered in connection with their blade alone counterparts. In fact, a recent investigation [8] has demonstrated that some wide chord fan blades exhibit a series of very closely spaced natural frequencies and that the corresponding blade alone mode shapes vary dramatically from blade to blade as a consequence of the small geometric and material differences between them.

In this context, it is desired here to assess the reliability of the RMS, ML, and IRMS identification techniques when two of the blade natural frequencies are close together. To this end, the two-degree-of-freedom shown in Fig. I.1 (model 2) was selected and the tuned values of the parameters k_2 and m_2 were varied to simulate different configurations. Of particular importance in the analysis of the veering phenomenon is the subsystem coupling which is here represented by the stiffness k_2. The reliability of the three identification methods will thus be assessed for different values of this parameter. In order to provide a worst case scenario, the mass m_2 was chosen, for a given stiffness k_2, so that the two natural frequencies of the system are as close as possible, i.e.,

$$m_2=\frac{k_2}{k_1-k_2}m_1. \tag{25}$$

Shown in Table 5 are the four sets of values of k_2 and m_2 that were considered as well as the corresponding relative differences in the frequencies ω_1 and ω_2, i.e., $2(\omega_2-\omega_1)/(\omega_2+\omega_1)$. For each of these four cases, the reliability of the forced response estimates was evaluated for different engine order excitations and it was found that the largest errors in the mean and standard deviation occurred at larger engine orders, $r=6$ to 9, only for case 4, see Fig. 5, while they were associated with $r=0$ for the three other blade models and for both the RMS and ML approaches. This observation which is somewhat contrary to prior findings, see

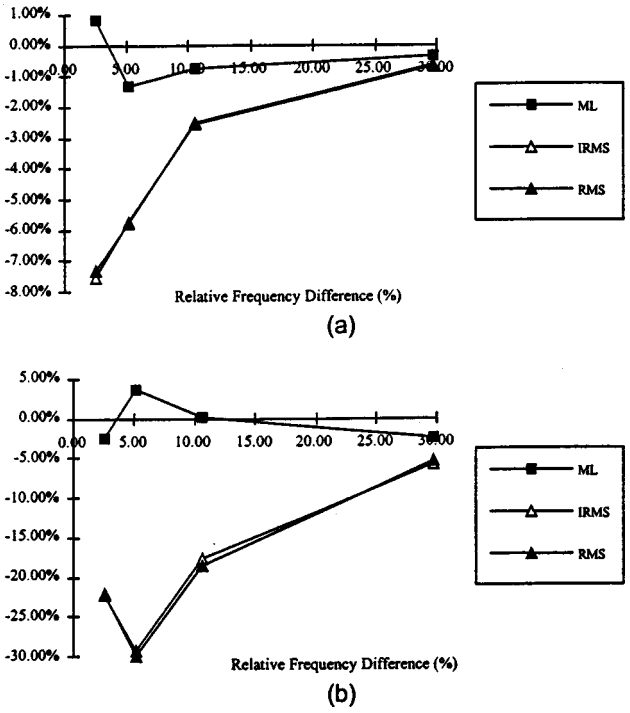


Fig. 7 Errors (*a*) in the mean and (*b*) in the standard deviation of the maximum response of the mass m_2 as functions of the relative difference of frequencies. ML, RMS and IRMS strategies, two-degree-of-freedom blade model 2 excited at the lowest tuned natural frequency corresponding to the zeroth engine order, $\omega_{0,1}^t$.

Fig. I.5 and 5 for example, may be explained by noting that the natural frequencies of the corresponding tuned disks exhibit a veering at $r=0$. Then, relying on past findings (see, for example, Kruse and Pierre [9] and Kaiser et al. [10]), it may be argued that the amplification of the forced response by mistuning would be particularly significant for such an excitation thereby increasing the likelihood of larger errors.

Consequently, shown in Fig. 7 are the relative errors in the mean and standard deviation of the maximum amplitude on the disk of the forced response of mass 2 for a zeroth-order excitation. Most notable on these figures are the sharp increases in the RMS errors as the two blade alone frequencies approach each other and the much less serious worsening of the ML forced response prediction, the latter of which is consistent with the continuous decrease of the mode shape-natural frequency coefficient of correlation as shown in Table 5. The improved behavior of the ML method over the RMS approach, although less dramatic, was also observed for other values of the engine order (see Delor [6]). Note finally that the iterative process described in connection with the satisfaction of the nonlinear frequency constraints in the ML formulation was found to be required in cases 1–3 as can be expected since an increased closeness of the two frequencies implies a more limited validity of the linearized approximation given by Eqs. (I.13) and (I.14).

The reliability of the IRMS method is not adequately described by Fig. 7 since this approach automatically reduces to the RMS technique for $r=0$. For other values of the engine order, it was typically found that the IRMS approach yielded errors of the same order of magnitudes as the ML method for cases 3 and 4 but exhibited substantially larger errors for case 2 and, especially, case 1. In conjunction with this observation, it was also found for these last two systems that Eqs. (23) and (24) did not yield acceptable values of Λ_{11} and Λ_{22}, i.e., real and positive, for a fraction of the blades analyzed that increased from 0 for $r=0$ to

almost 50 percent for case 1 with $r=12$. This absence of an IRMS model for certain sets of measured natural frequencies was proved [6] to be indeed associated with the closeness of the two natural frequencies and is by itself indicative of the breakdown of the logic associated with the formulation of the IRMS strategy. In this light, the appearance of large errors in the forced response prediction for the blades for which a model was in fact obtained is not surprising. On this basis, it is tentatively suggested that the IRMS method not be used when two natural frequencies are within 5 to 10 percent or less of each other.

Summary

This two-part investigation has focused on the estimation of the blade-to-blade variations of their structural properties, in particular masses and stiffnesses, from measurements of the natural frequencies of these blades only. Practically, this identification task does not represent an end result but rather a necessary first step in the accurate prediction of the forced response of turbomachine disks that would support the blades tested. It is thus in terms of forced response accuracy that the reliability of the suggested identification strategies has mostly been assessed.

In this second part of the investigation, a detailed study of the reliability and properties of the maximum likelihood (ML) formulation, developed in Part I, was first undertaken. This study revealed the following.

(i) The ML estimation strategy is most reliable when the assumed distribution of the parameters of the blades is accurate. However, even when this a priori model is vastly in error, i.e., the probability density function of the masses and stiffnesses is highly non-Gaussian (uniform for example) or the estimated covariance properties are far from their exact values, the estimates of the characteristics of the forced response still represent close approximations of the behavior of the disk, generally much better than the RMS prediction, see Tables 1 and 2 and Fig. 1.

(ii) The ML estimation approach is only mildly sensitive to small errors on the observed values of the blade natural frequencies. The reliability of the estimates of the forced response of the mistuned disks is unaffected by measurement noise of magnitude less than 1/4 to 1/3 of the variability of the natural frequencies corresponding to mistuning, see Table 3 and Fig. 2.

(iii) The determination of the entire set of natural frequencies of the blades is not necessary to obtain an accurate estimate of the forced response of the corresponding mistuned bladed disks; it is really only necessary to determine the natural frequencies of the blade modes (first bending, first torsion, etc.) that will be excited under the specified external loading, see Table 4.

(iv) The variability of the blade alone mode shapes can be quite well predicted by the ML estimation strategy although no direct measurement of these modes is used, see Figs. 3 and 4. This capability was found to be linked to the existence of a statistical correlation between the variability in blade alone mode shapes and natural frequencies. Further, an approximate expression for the corresponding coefficient of correlation was derived that was shown to be related directly to the reliability of the ML forced response prediction. Interestingly, it was shown that this coefficient could also be estimated from the population of identified blade models. Practically speaking, it is then possible to gauge directly, without any prior knowledge of the mistuning actually present, the reliability of the forced response statistics produced from the measured blade alone natural frequencies.

(v) Improvements in the reliability of the ML method in predicting the forced response can be achieved by biasing this technique toward higher mistuning levels. This was accomplished here by forcing the blade masses and stiffnesses to be close not to their tuned counterparts but rather to randomly perturbed (target) values thereof. This approach was shown to yield a reduction in the errors of the ML forced response prediction which is especially dramatic when the statistics of the target masses and stiffnesses closely match their exact mistuning counterparts.

It was demonstrated in the first part of this investigation that given the mass and mode shape mistuning of the blades, their modal stiffnesses can be selected to provide either accurate estimates of the blade alone natural frequencies or of the natural frequencies of the entire bladed disk, but not of both. Thus, by matching precisely the measured values of the blade alone frequencies, the RMS approach does not predict properly the behavior of the blades as components of the disk. This important observation suggested the introduction of a third technique, termed the improved random modal stiffnesses (IRMS) method, in which the random modal stiffnesses are in fact selected to match the behavior of the blade as part of a tuned assembly of similar blades. This approach is not per se an identification technique since it requires the prior determination of the ML estimates but it can be construed as a model simplification strategy or as a bridge between the ML and RMS formulations. As expected, the IRMS formulation led to errors in the forced response prediction that were smaller than their RMS counterparts although both of these approaches are similar in that they do not include mistuning in either mass or mode shape.

The last focus of the present investigation was on the assessment of the reliability of the three, ML, RMS, and IRMS, approaches when some of the blade alone frequencies are close together. This numerical study demonstrated an increase of the errors of the ensuing forced response prediction for all three methods but confirmed again the superiority of the ML formulation over its RMS counterpart.

Acknowledgments

The financial support of this investigation by a contract from the General Electric Aircraft Engine Company is gratefully acknowledged. The authors also wish to thank Dr. R. E. Kielb for insightful discussions of this problem.

References

[1] Collins, J. D., and Thomson, W. T., 1969, ''The Eigenvalue Problem for Structural Systems With Statistical Properties,'' AIAA J., **7**, pp. 642–648.
[2] Shinozuka, M., and Astill, C. J., 1972, ''Random Eigenvalue Problems in Structural Analysis,'' AIAA J., **10**, pp. 456–462.
[3] Vaicaitis, R., 1974, ''Free Vibrations of Beams with Random Characteristics,'' J. Sound Vib., **35**, pp. 13–21.
[4] Castanier, M. P., Ottarson, G., and Pierre, C., 1997, ''A Reduced Order Modeling Technique for Mistuned Bladed Disks,'' ASME J. Vibr. Acoust., **119**, pp. 439–447.
[5] Yang, M. T., and Griffin, J. H., 1997, ''A Reduced Order Approach for the Vibration of Mistuned Bladed Disk Assemblies,'' ASME J. Eng. Gas Turbines Power, **119**, pp. 161–167.
[6] Delor, J. P., 1998, ''Formulation and Assessment of Identification Strategies of Mistuned Bladed Disk Models from Free Response Data,'' M. S. thesis, Arizona State University.
[7] Wei, S. T., and Pierre, C., 1988, ''Localization Phenomena in Mistuned Assemblies With Cyclic Symmetry—Part I: Free Vibrations,'' ASME J. Vib. Acoust. Stress Reliability Design, **110**, pp. 429–438.
[8] Griffin, J. H., 1997, personal communication.
[9] Kruse, M. J., and Pierre, C., 1997, ''An Experimental Investigation of Vibration Localization in Bladed Disks. Part II. Forced Response,'' Presented at the 42nd International Gas Turbine and Aeroengine Congress and Exposition, Orlando, FL, ASME Paper 97-GT-502.
[10] Kaiser, T., Hansen, R. S., Nguyen, N., Hampton, R. W., Muzzio, D., Chargin, M. K., Guist, R., Hamm, K., and Walker, L., 1994, ''Experimental/Analytical Approach to Understanding Mistuning in a Transonic Wind Tunnel Compressor,'' NASA Tech. Memo., No. 108833, pp. 1–13.

Engine Fuel Droplet High-Pressure Vaporization Modeling

G.-S. Zhu

R. D. Reitz

Engine Research Center
University of Wisconsin—Madison,
1500 Engineering Drive, Madison, WI 53706

The objective of this investigation was to characterize the high-pressure vaporization processes of engine fuels, which are too complex in composition to be described with conventional methods. To do so a comprehensive model was developed for the transient vaporization process of droplets of practical engine fuels using continuous thermodynamics in which high-pressure effects are fully considered. Transport equations are derived in a spherical coordinate system for the semi-continuous systems of both gas and liquid phases. A general treatment of vapor-liquid equilibrium is presented, which can be applied with any type of cubic equation of state. Relations for the properties of the continuous species are formulated. The model is further applied to calculate the transient high-pressure vaporization processes of droplets of representative engine fuels—diesel and gasoline. The high-pressure vaporization processes of droplets of two single-component fuels are also predicted for comparison. The results clarify the characteristics of the vaporization processes of engine fuel droplets and indicate the significant effects of fuel type on the vaporization behavior. The comparison with the results of the single-component cases also emphasizes the importance of considering the influence of multi-component fuels in practical applications. [DOI: 10.1115/1.1361058]

Introduction

Droplet high-pressure vaporization is of special interest to the research of internal combustion engines due to the high-pressure conditions in the combustion chamber when the fuel droplets are sprayed. The resulting droplet vaporization process typically has two important characteristics. One is the process happens in a high-pressure environment, causing many important aspects that are not adequately considered by low-pressure models such as the gas absorption into liquid phase, the pressure-dependent thermophysical properties, the more complex vapor-liquid equilibrium, and the possible critical state at the droplet surface. Another characteristic is that the fuel droplet is usually complex and composed of hundreds of species. However, these characteristics are not considered by most of vaporization models currently employed for engine theoretical analyses.

Several investigations have been conducted on droplet high-pressure vaporization, including the studies reviewed by Givler and Abraham [1] and those reported recently by Stengele et al. [2], Curtis et al. [3], Aggarwal et al. [4], Zhu and Aggarwal [5,6], and Givler and Abraham [7]. Most of these works concentrated on the vaporization of single-component fuel droplets. Few works appear in the literature on liquid mixtures, in which the liquid mixtures were represented by just two components with distinctly different boiling temperatures [e.g., Lazar and Faeth [8], Aggarwal, et al. [4], Stengele et al. [2], and Hsieh et al. [9]]. In order to reproduce the distillation characteristics of practical engine fuels with acceptable accuracy, the number of discrete components chosen to represent practical fuels should be large enough. As a result, a high requirement of computer resources is needed to perform numerical analyses with these methods.

The multicomponent nature of engine fuel composition plays an important role on the engine performance. The flame propagation in engine combustion chambers is a function of both the fuel composition and the mixture ratio (Kuo [10] and Turns [11]) and thus is directly linked to the engine performance. The spatial distribution of the mixture ratio and the fuel composition are also important considerations for determining the location of the spark plug. The volatility of the fuel has a significant effect on diesel spray penetration (Robert et al. [12]) and thus influences the diesel fuel time-space distribution. For port injection gasoline engines, the fuel composition and the corresponding volatility are essential to the engine transient behavior (Shayler et al. [13]). The fuel chemical reactions and the pollutant formation and oxidation processes are directly related to the fuel composition. These issues highlight the necessity of considering fuel composition in engine fuel modeling.

A new approach to treating real fuel mixtures is to employ the thermodynamics of continuous systems, in which the fuel composition, and consequently the system properties, are represented and described by a continuous probability density function. This idea of treating complex mixtures has been mostly applied in chemical engineering. Only recently was it applied to the field of droplet evaporation by Tamim and Hallett [14] and to the field of spray combustion by Lippert and Reitz [15] and Lippert [16]. These studies have made important steps in the low-pressure vaporization applications of continuous thermodynamics, however, there is generally no model available for treating the high-pressure vaporization processes of droplets of practical liquid fuels. In this paper, a comprehensive model is developed for the high-pressure vaporization processes of complex fuel droplets using continuous thermodynamics. Typical results of representative engine fuels—diesel and gasoline—along with the results for two single-component fuels for comparison, are then presented and analyzed.

Problem Formulation

The formulation considered here deals with two semi-continuous mixture systems, which include both discrete species (such as the nitrogen for the gas phase and the dissolved gas for the liquid phase) and continuous ones (such as the fuel mixture vapor for the gas phase and the liquid fuel for the liquid phase). A general molar distribution function for the compositions of the systems is defined as

Contributed by the Internal Combustion Engine Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the ICE Division July 11, 2000; final revision received by the ASME Headquarters Dec. 6, 2000. Associate Editor: D. Assanis.

$$G^p(\tau)=y_f{}^p f^p(\tau)+\sum_{s=1}^{N} y_s{}^p \delta(\tau-\tau_s) \tag{1}$$

$$\int_0^\infty G^p(\tau)d\tau=1, \quad \int_0^\infty f^p(\tau)d\tau=1, \quad \sum_{s=1}^{N} y_s{}^p=1-y_f{}^p \tag{2}$$

where the superscript p represents v or l, denoting the properties of the gas or liquid phase, respectively. y_s is the mole fraction of the sth discrete species, y_f the mole fraction of the continuous one, N the total number of discrete species, and δ the dirac delta function. τ is some characterizing property that is chosen to be molecular weight in this paper. The molar distribution function $f(\tau)$ is only applicable for the continuous species. In the present paper, the Γ distribution was chosen for the distributions of both fuel liquid and vapor phases such that

$$f(\tau)=\frac{(\tau-\gamma)^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)}\exp\left[-\frac{(\tau-\gamma)}{\beta}\right] \tag{3}$$

where γ is the origin, α and β are parameters controlling the shape, and $\Gamma(\alpha)$ is the gamma function, with the mean molecular weight $\theta=\alpha\beta+\gamma$ and variance $\sigma^2=\alpha\beta^2$. With the use of the continuous thermodynamics theory (Gal-Or et al. [17] and Zhu and Reitz [18]) for the continuous species, the conventional theory for the discrete species, and assumptions that fuel droplets vaporize in nonconvective stagnant environments, we derive in the following, in a spherical coordinate system, the transient governing equations for both gas and liquid phase semicontinuous systems, along with the interphase conditions at the droplet surface.

Governing Equations for Gas Phase. The continuity equation for the N discrete species is given by

$$\frac{\partial \rho_s}{\partial t}+\frac{1}{r^2}\frac{\partial}{\partial r}(\rho_s u r^2)=\frac{1}{r^2}\frac{\partial}{\partial r}\left[\rho D_{sm} r^2 \frac{\partial}{\partial r}\left(\frac{\rho_s}{\rho}\right)\right] \quad (s=1,2,\ldots N) \tag{4}$$

where u is the molar-averaged velocity, ρ_s and ρ are the molar densities of the sth discrete component and the total mixture, respectively. D_{sm} is the diffusivity of the sth species with respect to the gas mixture. Here Fick's law is assumed to be applicable for the semicontinuous system. The continuity equation for the continuous species can be derived as (Zhu and Reitz [18])

$$\frac{\partial \rho y_f^v}{\partial t}+\frac{1}{r^2}\frac{\partial}{\partial r}(\rho y_f^v u r^2)=\frac{1}{r^2}\frac{\partial}{\partial r}\left[\rho \overline{D}_m r^2 \frac{\partial y_f^v}{\partial r}+y_f^v \frac{\partial}{\partial r}(\rho \overline{D}_m r^2) - y_f^v \int_0^\infty f^v(\tau)\frac{\partial}{\partial r}(\rho D_{\tau m} r^2)\,d\tau\right] \tag{5}$$

$$\frac{\partial}{\partial t}[\rho y_f^v (\theta^n)^v]+\frac{1}{r^2}\frac{\partial}{\partial r}[\rho y_f^v (\theta^n)^v u r^2] = \frac{1}{r^2}\frac{\partial}{\partial r}\left\{\rho D_m^{\theta n} r^2 \frac{\partial}{\partial r}[y_f^v (\theta^n)^v]+y_f^v (\theta^n)^v \frac{\partial}{\partial r}(\rho D_m^{\theta n} r^2) - y_f^v \int_0^\infty f^v(\tau)\tau^n \frac{\partial}{\partial r}(\rho D_{\tau m} r^2)d\tau\right\} \quad (n=1,2,\ldots) \tag{6}$$

where

$$\overline{D}_m=\int_0^\infty D_{\tau m} f^v(\tau)\,d\tau$$

$$D_m{}^{\theta n}(\theta^n)^v=\int_0^\infty D_{\tau m}\tau^n f^v(\tau)\,d\tau \quad (n=1,2,\ldots)$$

where $(\theta^n)^p$ is the nth moments about the origin of the vapor or liquid distribution, which is defined as

$$(\theta^n)^p=\int_0^\infty \tau^n f^p(\tau)d\tau, \quad (n=1,2,\ldots) \tag{7}$$

Any number of equations can be derived from Eq. (6) with different values of n, but for the two-parameter distribution of Eq. (3), $n=2$ will suffice for the calculations.

The momentum equation for the gas mixture is given by

$$\frac{\partial \rho u}{\partial t}+\frac{1}{r^2}\frac{\partial}{\partial r}(\rho u^2 r^2)+\frac{\partial P}{\partial r}=\frac{1}{r^2}\frac{\partial}{\partial r}\left[2\mu r^2\frac{\partial u}{\partial r}+\lambda\frac{\partial}{\partial r}(ur^2)\right] - \frac{2}{r^2}\left[2\mu u+\frac{\lambda}{r}\frac{\partial}{\partial r}(ur^2)\right] \tag{8}$$

where P is pressure, μ and λ are the viscosity and the second viscosity coefficient, respectively. For a nonreacting system, the energy equation for the mixture is

$$\frac{\partial \rho I}{\partial t}+\frac{1}{r^2}\frac{\partial}{\partial r}(\rho I u r^2)+\frac{P}{r^2}\frac{\partial}{\partial r}(ur^2) = \frac{1}{r^2}\frac{\partial}{\partial r}(r^2 J_h)+\frac{\partial u}{\partial r}\left[2\mu\frac{\partial u}{\partial r}+\frac{\lambda}{r^2}\frac{\partial}{\partial r}(ur^2)\right] + \frac{2u}{r^2}\left[2\mu u+\frac{\lambda}{r}(ur^2)\right] \tag{9}$$

where I is the specific internal energy of the gas mixture, which is related to the pressure and temperature by

$$I=\left[\sum_{s=1}^{N}\frac{\rho_s}{\rho}h_s(T)+\int_0^\infty f(\tau)h(\tau,T)\,d\tau\right]-\frac{P}{\rho} \tag{10}$$

where h is the enthalpy, which is obtained from JANAF tables for the discrete species and is obtained for the continuous fuel species from the following formula

$$h(\tau,T)=[c_{p1}(T)+c_{p2}(T)\tau]T \tag{11}$$

The coefficients c_{p1} and c_{p2} are given by Chou and Prausnitz [19] for various fuels. The energy flux J_h in Eq. (9) is given by

$$J_h=k\frac{\partial T}{\partial r}+\rho\left\{\sum_{s=1}^{N} D_{sm}h_s(T)\frac{\partial}{\partial r}\left(\frac{\rho_s}{\rho}\right)+D_m{}^h\overline{h}(T)\frac{\partial y_f^v}{\partial r} + y_f^v\frac{\partial}{\partial r}[D_m{}^h\overline{h}(T)]-y_f^v\int_0^\infty f^v(\tau)\frac{\partial}{\partial r}[D_{\tau m}h(\tau,T)]\,d\tau\right\} \tag{12}$$

where k is the thermal conductivity of the mixture, and

$$D_m{}^h\overline{h}(T)=\int_0^\infty f^v(\tau)D_{\tau m}h(\tau,T)\,d\tau \tag{13}$$

The cubic equation of state (EOS) is employed for the semicontinuous system and the subsequent vapor-liquid equilibrium calculation, which can be written in a general form as (Reid et al., [20])

$$P=\frac{R_u T}{V-b}-\frac{a}{V^2+qbV+wb^2} \tag{14}$$

where q and w are constants depending on the type of cubic equation of state. a and b are functions of temperature and species mole fraction. The formulas in Zhu and Reitz ([18]) are employed for calculating the constants a and b for the semicontinuous systems containing hydrocarbon fuels.

Governing Equations for Liquid-Phase. The forms of the governing equations used for the liquid-phase directly relate to the model employed to describe the droplet interior flow. As indicated in Tamim and Hallett [14], the effects of departures from the

well-mixed state are of less importance for mixtures with closely spaced component boiling points, and further, the effects on the total fuel concentration are even less. Studies of high-pressure evaporation have also shown that the gas absorption is mainly limited in a thin outer layer of the vaporizing droplet throughout the droplet lifetime (Givler and Abraham [7] and Zhu and Aggarwal [6]). Here we approximate the droplet interior with a two-zone model, in which the dissolved gas is assumed to exist only in the outer layer of the droplet and all other liquid properties are assumed well mixed. Similar models have been employed by other researchers (Curtis et al. [3] and Gradinger and Boulouchos [21]). With this interior model and based on the mass balance on the droplet surface, the governing equations for the mass and the composition of the liquid phase can be derived as (Zhu and Reitz [18])

$$(x_{feq}-y_f^v)|_{R^+}\dot{m}=c_l R\frac{v_a}{3}\frac{dx_f}{dt}-\left[\rho\overline{D}_m\frac{\partial y_f^v}{\partial r}+y_f{}^v\frac{\partial}{\partial r}(\rho\overline{D}_m)-y_f^v\int_0^\infty f^v(\tau)\frac{\partial}{\partial r}(\rho D_{\tau m})\,d\tau\right]\Bigg|_{R^+} \tag{15}$$

$$\frac{c_l R}{3}x_{feq}\frac{d(\theta^n)^l}{dt}=\Bigg\{[x_{feq}(\theta^n)^l-y_f{}^v(\theta^n)^v]\dot{m}-\frac{c_l R}{3}v_a(\theta^n)^l\frac{dx_f}{dt}+\rho D_m{}^{\theta n}\frac{\partial}{\partial r}[y_f^v(\theta^n)^v]+y_f(\theta^n)^v\frac{\partial}{\partial r}(\rho D_m{}^{\theta n})-y_f{}^v\int_0^\infty f^v(\tau)\tau^n\frac{\partial}{\partial r}(\rho D_{\tau m})d\tau\Bigg\}\Bigg|_{R^+}$$

$$(n=1,2,\ldots) \tag{16}$$

where $\dot{m}$ is the vaporization rate. c_l and x_f are the molar density and the fuel molar fraction of the liquid phase, respectively. R is the droplet radius. $(\theta^n)^l$ is the nth moments about the origin of the liquid distribution and is defined in Eq. (7). The equivalent molar fraction of liquid phase x_{feq} is defined as

$$x_{feq}=v_a x_f+(1-v_a) \tag{17}$$

where v_a is the ratio of the volume of the absorbed-gas layer to the total droplet volume. Zero and unity values of v_a correspond to no gas absorption and the infinite diffusion coefficient condition, respectively. However, the present calculations show that v_a has little effect on the droplet vaporization behavior except for a slight influence on the results of the liquid composition. In the following, v_a is taken to be 0.5.

Boundary Conditions and Vapor-Liquid Equilibrium. At the droplet surface, the boundary conditions are expressed using the mass conservation Eq. (15), the energy Eq. (18), and thermodynamic equilibrium, Eq. (19):

$$\frac{R}{3}c_l c_{pl}\frac{\partial T_l}{\partial t}=-k\frac{\partial T}{\partial r}\Bigg|_{R^+}+\sum_{s=1}^{N}J_s\Delta H_{vs}+\int_0^\infty J_\tau\Delta H_{v\tau} \tag{18}$$

$$G^v(\tau)\varphi^v(\tau)=G^l(\tau)\varphi^l(\tau) \tag{19}$$

where ΔH_{vs} and $\Delta H_{v\tau}$ are the enthalpies of vaporization of the discrete and the continuous species, respectively. J_s and J_τ are the molar fluxes on the droplet surface of the discrete and continuous components, respectively. They are given by

$$\left[J_s=\dot{m}y_s-\rho D_{sm}\frac{\partial y_s}{\partial r}\right]\Bigg|_{R^+} \tag{20}$$

$$J_\tau=\left[\dot{m}y_f^v-\rho\overline{D}_m\frac{\partial y_f^v}{\partial r}-y_f^v\frac{\partial}{\partial r}(\rho\overline{D}_m)+\int_0^\infty f^v(\tau)\frac{\partial}{\partial r}(\rho D_{\tau m})d\tau\right]\Bigg|_{R^+} \tag{21}$$

Equation (19) expresses the vapor-liquid equilibrium at the droplet surface. For the semicontinuous system considered here and using cubic EOS (14), the fugacity coefficients are derived as

$$\ln\varphi(k)=\frac{b(k)}{b}(z-1)-\ln(z-B^*)+A_{\mathrm{EOS}}\left(\frac{b(k)}{b}-C_k\right) \tag{22}$$

where

$$A_{\mathrm{EOS}}=\frac{A^*}{B^*\sqrt{q^2-4w}}\ln\frac{2z+B^*(q+\sqrt{q^2-4w})}{2z+B^*(q-\sqrt{q^2-4w})}$$

$$A^*=\frac{aP}{R_u{}^2T^2},\quad B^*=\frac{bP}{R_uT}$$

For discrete species, $k=s$, and

$$C_s=\frac{2}{a}\left[\sum_{j=1}^{N}y_j^p a_{sj}+y_f^p\int_0^\infty f^p(\tau)a_{s\tau}\,d\tau\right] \tag{22a}$$

For continuous species, $k=\tau$, and

$$C_\tau=\frac{2}{a}\left[\sum_{j=1}^{N}y_j^p a_{\tau j}+y_f^p\int_0^\infty f^p(\tau)a_{\tau\tau}\,d\tau\right] \tag{22b}$$

where the quantities a with double subscripts are cross terms calculated via mixing rule (Reid et al., [20]). With Eq. (22) and the formulas for EOS parameters a and b given by Zhu and Reitz [18], the ratio of the fugacity coefficients in the vapor and liquid phases for the continuous fractions are further expressed as an explicit function of molecular weight

$$\ln\varphi(\tau)=C_k{}^{(1)}+C_k{}^{(2)}\tau \tag{23}$$

where the coefficients $C_k{}^{(1)}$ and $C_k{}^{(2)}$ are derived as

$$C_k^{(1)}=\frac{b_0}{b}(z-1)-\ln(z-B^*)+A_{\mathrm{EOS}}\left\{\frac{b_0}{b}-\frac{2a_0}{a}\left[\sum_{j=1}^{N}y_j a^{0.5}(j,j)\times(1-k_{\tau j})+y_f(a_0+a_1\theta)\right]\right\} \tag{24}$$

$$C_k^{(2)}=\frac{b_1}{b}(z-1)+A_{\mathrm{EOS}}\left\{\frac{b_1}{b}-\frac{2a_1}{a}\left[\sum_{j=1}^{N}y_j a^{0.5}(j,j)\times(1-k_{\tau j})y_f(a_0+a_1\theta)\right]\right\} \tag{25}$$

where the quantities k with double subscripts are interaction coefficients, which are taken to be zeros in this paper. It is worth noting that the equations derived above are general for a cubic EOS. Therefore, any type of cubic EOS covered by Eq. (14) can conveniently be used for vapor-liquid equilibrium and/or vaporization process calculations using the above equations. In this paper the Peng-Robinson EOS was selected for both the vapor-liquid equilibrium and the vaporization process calculations.

For a mixture system, the heat of vaporization of each species is defined as the difference between the partial molar enthalpy of that species in the vapor and liquid phases. The following thermodynamic relation gives the partial molar enthalpy for both the discrete and continuous species

$$\overline{H}_k-\overline{H}_k^0=-R_uT^2\frac{\partial}{\partial T}(\ln\varphi^k) \tag{26}$$

Thermophysical Properties. The thermodynamic and transport properties are considered to be functions of pressure, temperature, and composition. A set of empirical formulas is derived for calculating the critical properties and the boiling temperature of the continuous species. They are formulated for n-alkanes from c_2 to c_{20} with the following form:

Table 1 Constants for calculating critical properties and boiling temperature for alkane fuels c_2 to c_{20}

	Critical pressure (bar)	Critical temperature (K)	Boiling temperature (K)
χ_0	5.5590×10^1	2.2636×10^2	1.1166×10^2
χ_1	-3.3177×10^{-1}	3.6471×10^0	2.9582×10^0
χ_2	6.3261×10^{-4}	-6.3057×10^{-3}	-4.2393×10^{-3}

$$\chi = \chi_0 + \chi_1 \tau + \chi_2 \tau^2 \tag{27}$$

where τ is the molecular weight. Table 1 gives all the constants for the specified range of molecular weights. It was found that the relative errors of these formulas are generally less than 3 percent as compared to the data of Reid [20].

The method reported by Chung et al. [22] was employed to calculate the thermal conductivity and viscosity of the mixture at high pressures. For discrete species, the diffusivity was calculated by means of the Chapman-Enskog theory in conjunction with the collision integrals given by Reid et al. [20]. For continuous species, the formulas of Tamim and Hallett [14] were employed for the diffusivities of the vapor, and the first and second moments of the composition. The diffusivity of the enthalpy diffusion in Eq. (13) was taken to be the same as that of fuel vapor. All the diffusivities were corrected for pressure effects by using the Takahashi correlation (Reid et al. [20]). A generalized thermodynamic correlation based on a three-parameter corresponding state (Byung and Michael [23]) was used to calculate the enthalpy correction for high-pressure effects. The liquid density was directly calculated by Peng-Robinson EOS.

An arbitrary Lagrangian-Eulerian numerical method with a dynamically adaptive mesh was employed to solve the governing equations for the gas-phase system (Zhu and Aggarwal [6]). The governing equations for the liquid-phase system were solved using a Runge-Kutta numerical method. The solutions for both systems are coupled with each other through the regressing droplet surface conditions.

Results and Discussion

Two representative engine fuels—diesel and gasoline—were chosen to be the liquid mixtures studied. The composition parameters of the diesel fuel were: $\alpha_L = 18.5$, $\beta_L = 10.0$, and $\gamma_L = 0.0$, with $\theta_L = 185$ and $\sigma_L = 43$, and those of the gasoline fuel were $\alpha_L = 5.7$, $\beta_L = 15.0$, and $\gamma_L = 0.0$, with $\theta_L = 85.5$ and $\sigma_L = 35.8$ (Tamim and Hallett [14]). These parameters were chosen so that the calculated distillation curves with the distributions reproduce best the corresponding ASTM D86 experimental data. Very narrow distributions were also employed to approximate the single-component test liquid, with $\alpha_L = 100$, $\beta_L = 0.1$, and γ_L set so that the mean molecular weight of the narrow distribution was equal to the molecular weight of the single-component liquid to be represented. With this method, the single-component liquid is approximately *n*-tridecane for diesel and *n*-hexane for gasoline.

Calculations were first carried out for the vapor-liquid equilibrium of a system containing liquid and nitrogen. Figure 1 presents the phase-equilibrium calculations for single-component *n*-butane-nitrogen, *n*-heptane-nitrogen and *n*-decane-nitrogen systems, compared with the corresponding experimental data (Knapp et al. [24]) for each system. Each *n*-alkane was represented by a very narrow distribution by taking the parameters with the method stated above. For all three systems, the theoretical prediction matches the experimental data for the liquid-phase equilibrium mole fraction closely, but deviates slightly from the gas-phase data. This may be attributed to the intrinsic limitations of the cubic EOS employed and the precision of the formulas for calculating the EOS parameters *a* and *b*. Nevertheless, overall good agreement is obtained. This indicates the applicability of the present vapor-liquid equilibrium model for hydrocarbon liquid mixtures.

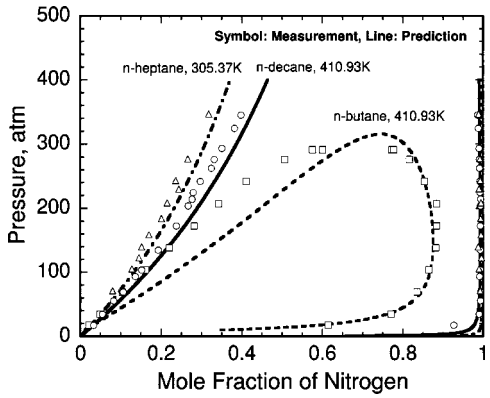


Fig. 1 Predictions of single-component *n*-butane-nitrogen, *n*-heptane-nitrogen, and *n*-decane-nitrogen vapor-liquid equilibrium systems using semicontinuous vapor-liquid equilibrium model, compared with corresponding experimental data of Knapp et al. [24]

Figures 2(*a*) and 2(*b*) show the phase-equilibrium concentrations versus temperature at four reduced pressures for both diesel-nitrogen and gasoline-nitrogen systems (solid curves), respectively, together with results for two single-component liquid-nitrogen systems (dashed curves). Here a single-component liquid with $\alpha_L = 100.0$, $\beta_L = 0.1$, and $\gamma_L = 175.0$ is used for comparison with diesel and one with $\alpha_L = 100.0$, $\beta_L = 0.1$, and $\gamma_L = 75.5$ is used for comparison with gasoline. The P_r represents the reduced pressure of the pure fuel. (The critical pressure used here for P_r is determined to be 18.6 atm for diesel and 33.0 atm for gasoline). The results of both liquid mixture-nitrogen systems generally show similar characteristics, e.g., with increasing pressures the critical mixing temperature and the fuel vapor mole fraction decrease, and the gas absorption in liquid phase increases and becomes significantly large in the supercritical regime. However, for the same reduced pressure, the gas absorption in liquid gasoline is obviously larger than that in liquid diesel, whereas the critical mixing temperature of the gasoline case is considerably smaller than that of the diesel case.

For the diesel case, as shown in Fig. 2(*a*), the predicted gas solubility is very close to that of its single-component counterpart case. Only at high temperature is the gas solubility of the single-component liquid slightly larger. However, the single-component liquid shows a higher boiling temperature and a smaller critical mixing temperature. At a certain temperature, the vapor mole fraction of the diesel case is larger than that of its single-component counterpart case over the whole subcritical pressure range or in the supercritical pressure range beyond the vicinity of the critical mixing point. For the gasoline case as shown in Fig. 2(*b*), the above-mentioned differences for the diesel case become more pronounced. Consequently, the single-component droplet vaporization process may occur with a very different vaporization rate and reach its critical state much earlier. These phenomena clearly display the difference in phase-equilibrium results for both kinds of fuels and imply a strong influence of fuel composition on the droplet vaporization process, especially for volatile fuels.

Figure 3 presents the critical mixing pressures versus critical mixing temperature for phase equilibria of both diesel- and gasoline-nitrogen systems. The critical mixing pressure decreases with increasing temperature. The critical mixing pressures of the diesel-nitrogen system are considerably larger than those of the gasoline-nitrogen system. Due to nitrogen being the dominant species in air, these results imply that, for typical engine conditions, it may be impossible for gasoline fuel in homogenous

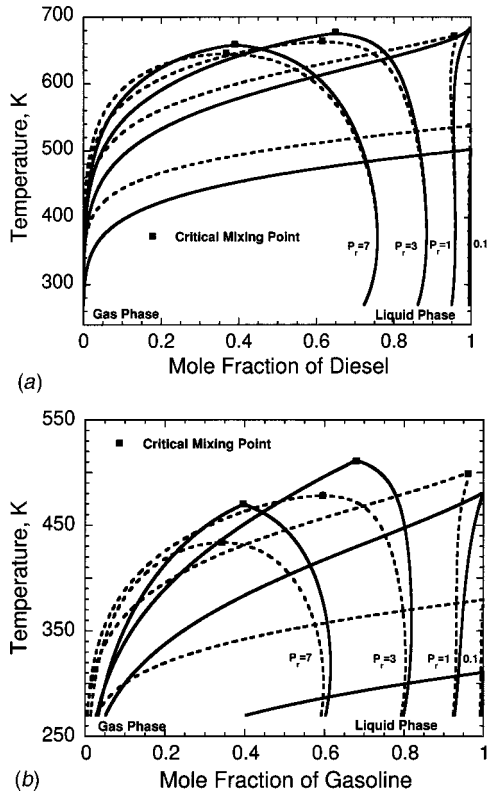


Fig. 2 Phase-equilibrium in terms of pressure-temperature diagrams for (*a*) diesel-nitrogen system, diesel fuel with distribution parameters $\alpha_L=18.5$, $\beta_L=10.0$, and $\gamma_L=0.0$, and for (*b*) gasoline-nitrogen system, gasoline fuel with parameters $\alpha_L=5.7$, $\beta_L=15.0$, and $\gamma_L=0.0$. Solid lines: real fuels; Dashed lines: single-component fuels.

charged gasoline engines to reach its critical mixing state, but it is possible for diesel fuel in diesel engines to experience a critical mixing vaporization process, see Fig. 3.

We now focus on the transient vaporization process of a droplet of diesel fuel and a droplet of gasoline fuel. The corresponding single-component liquid droplets were also calculated and are presented for comparison. The droplets are initially 0.03 mm in diameter and at 273 K and are suddenly put into a nitrogen environment at 950 K for the diesel droplet and at 500 K for the

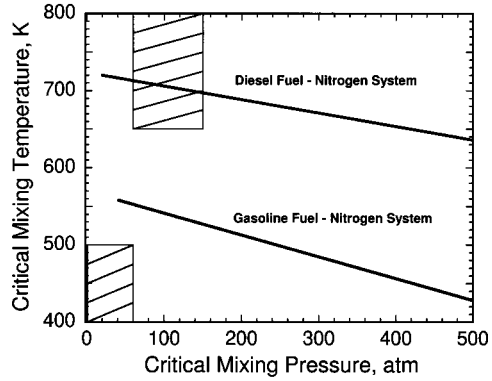


Fig. 3 Critical mixing temperature versus critical mixing pressure for both diesel-nitrogen and gasoline-nitrogen vapor-liquid equilibrium systems. Typical homogenous charge gasoline and diesel engine operating conditions are shown in the cross-hatched regions.

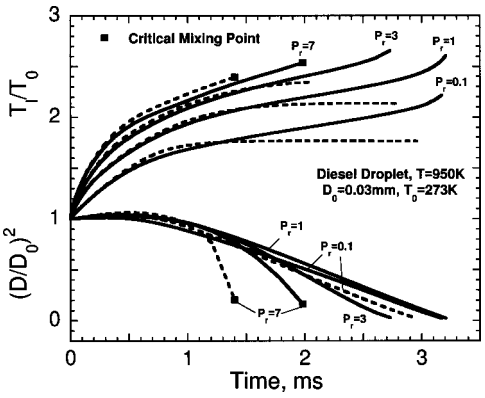


Fig. 4 Surface temperatures (upper) and area regression rates (lower) of droplets versus time at four different reduced pressures. Solid lines: diesel fuel. Dashed lines: single-component fuel.

gasoline droplet. These environmental parameters are approximately the conditions experienced by these fuels in engines (see Fig. 3). The environmental pressures considered ranged from sub- to supercritical.

Figure 4 (upper part) presents the calculated diesel droplet temperature versus time at four different reduced pressures, for both diesel fuel (solid curves) and its single-component fuel counterpart (dashed curves). An obvious characteristic is that the temperature of the liquid mixture droplet is entirely transient no matter whether the pressure is subcritical or supercritical, whereas the single-component droplet reaches a steady temperature at subcritical conditions. For the liquid mixture droplet, the fuel boiling temperature continuously increases with increasing droplet temperature. This is seen in Fig. 5, where both the variation in droplet boiling temperature, T_b at $P_r=0.1$ and the critical mixing temperature, T_c at $P_r=3.0$ are presented. As the evaporation process proceeds, the lighter species are continuously preferentially distilled, and the heavy species become more and more dominant in the composition (see Fig. 7). Correspondingly, the boiling and critical mixing temperatures increase continuously. The former increases almost at the same rate as that of the increase in droplet surface temperature. Therefore, the diesel droplet never reaches the boiling state, and the heat-up process spans the whole droplet lifetime. Due to the continuous increase in critical mixing temperature and the relatively small droplet size, the diesel droplet reaches its critical mixing state at a relatively high reduced pressure of $P_r=7$.

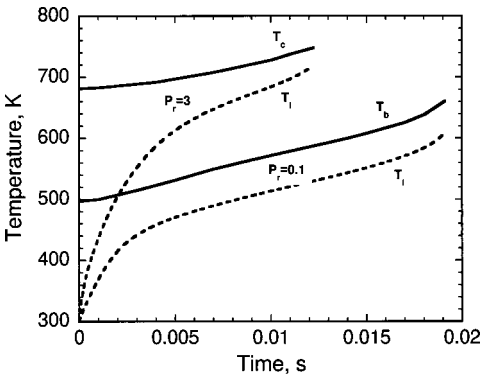


Fig. 5 Time histories of boiling temperature at reduced pressure of 0.1 and critical mixing temperature at reduced pressure of 3.0 for diesel droplet

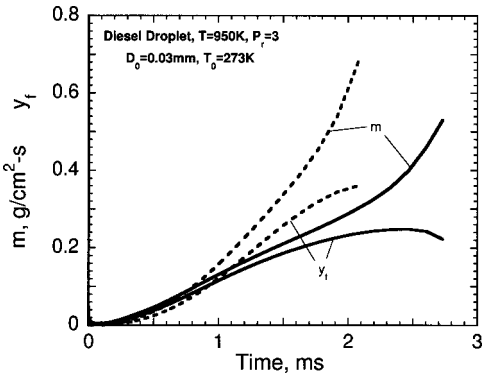


Fig. 6 Mass vaporization rate and vapor mole fraction at droplet surface versus time at reduced pressure of $P_r=3$. Solid lines: diesel fuel. Dashed lines: single-component fuel.

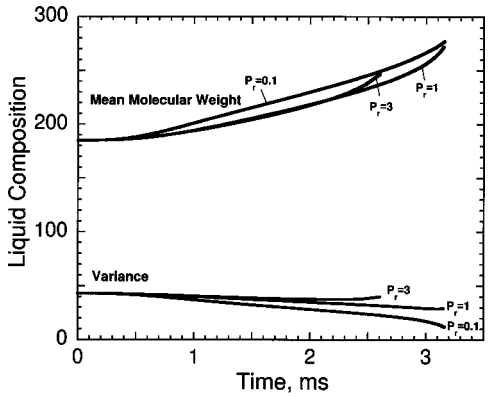


Fig. 7 Mean molecular weight of diesel droplet and variance of diesel liquid phase distribution versus time at three different reduced pressures

Figure 4 (lower part) also shows the change of the diesel droplet surface area versus time at the same four reduced pressures (solid lines), together with the results for the single-component liquid (dashed lines) at two pressures of $P_r=0.1$ and $P_r=7$. The droplet surface regression rate increases progressively with pressure mainly as a result of the reduced heat of vaporization and the reduced liquid density due to the rapid increase in gas absorption at high pressures. The remaining heavy components in the fuel mixture causes the regression rate to decrease near the end of its lifetime (e.g., see $P_r=3$ case). An interesting behavior on these curves is that volumetric dilation occurs during the initial period of vaporization. This is because the initial surface vapor fraction and the vaporization rate are very small due to the low volatility of the fuels. Consequently, the change in density effect dominates the initial regression process and contributes to the volumetric dilation. It is also worth noting that the single-component droplet has a larger vaporization rate and a shorter lifetime, and it gains its critical mixing state much earlier. Siebers [25] and Higgins et al. [26] have experimentally observed that the multicomponent fuels have longer liquid penetration lengths compared to single-component fuels. Lippert [16] has also numerically predicted that the amount a single-component spray mass vaporizes through critical phase change is larger than that of a multicomponent spray mass. These findings are totally supported by the results presented here.

Figure 6 presents the vaporization rate and the fuel vapor mole fraction at the droplet surface at a pressure of $P_r=3$. Compared to its single-component case, diesel fuel has a larger vapor mole fraction at the initial stage due to the lighter components distilling off first, and a smaller one at the later stages due to the heavier component remaining in liquid phase. Consequently, the vaporization rate of the diesel fuel droplet is initially larger and becomes smaller later on. As also seen from the figure, the vaporization rate and the surface vapor mole fraction of both fuels are all extremely small, especially at the beginning of the process which, as stated above, partly contributes to the obvious initial volumetric dilation. Similarly, the smaller initial vaporization rate and surface vapor mole fraction of the single-component fuel results in a slightly more obvious volumetric dilation, and their subsequent more rapid rise contributes to the larger surface regression rate, and hence a shorter lifetime, as shown in Fig. 4.

Figure 7 shows the time histories of the mean molecular weight and the width of the composition distribution of the diesel liquid for the same three cases as those of Fig. 4. As the vaporization process proceeds, the volatile components continuously distill out, resulting in a continuous increase in mean molecular weight of the liquid. Correspondingly, the width of the liquid distribution decreases as the lighter end of the distribution shrinks.

Figure 8 presents the change of the nondimensional droplet surface area versus time at three different reduced pressures, for both the gasoline droplet (solid curves) and its single-component counterpart (dashed curves). Due to the high volatility of this fuel, no volumetric dilation exists during the initial vaporization period. However, a more pronounced reduced regression rate is observed for gasoline fuel during the late period of vaporization, especially at higher pressures. Consequently, the gasoline droplet has a considerably larger droplet lifetime than its single-component counterpart does. This point can also be seen in Fig. 9, where the droplet lifetimes of both diesel and gasoline (solid curves) and

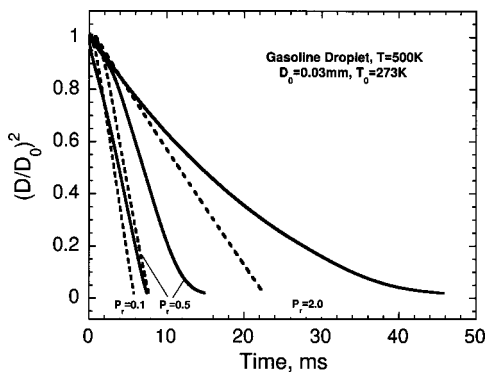


Fig. 8 Surface area regression rates of droplets versus time at three different reduced pressures. Solid lines: gasoline fuel. Dashed lines: single-component fuel.

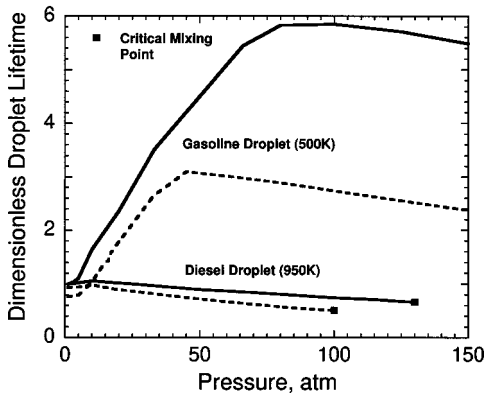


Fig. 9 Normalized lifetimes versus ambient pressure of droplets of diesel and gasoline and the single-component counterpart fuels. Solid lines: real fuels. Dashed lines: single-component fuels.

their single-component counterparts (dashed curves) are presented. The lifetimes are normalized by the lifetime of the diesel droplet at 1 atm (3.14 ms) for the results of diesel and its single-component counterpart, and by the lifetime of the gasoline droplet at 1 atm (7.40 ms) for those of gasoline and its single-component counterpart. For single-component droplet, the change of lifetime with increased pressure is a competing result of the increased droplet heat-up time, reduced mass transfer rate, and reduced heat of vaporization. But for the liquid mixture droplets, besides the above competing factors, the lifetime is also closely related to the characteristics of the fuel composition. As seen from the figure, the lifetime of the diesel droplet increases with increasing pressure at low pressures and then decreases with increasing pressures. The single-component counterpart droplet has a shorter lifetime and the difference between them increases with increasing pressure. However, for gasoline and its single-component counterpart, the lifetimes increase drastically with increasing pressure in a pressure range that the fuel may experience in gasoline engines (see Fig. 3). At higher pressures, a huge difference occurs between the predicted lifetimes of both fuel droplets. Nevertheless, the general trends with increasing pressure of both the mixture fuels are similar to those of their single-component counterparts.

Conclusions

The transient high-pressure vaporization processes of practical engine fuel droplets were investigated. A comprehensive model has been developed for droplets of complex liquid mixtures using continuous thermodynamics, which considers the transient, spherically-symmetric conservation equations for both the gas and liquid semicontinuous phases, pressure-temperature-composition variable thermophysical properties, and a general treatment of the vapor-liquid equilibrium at the droplet surface. This model provides a practical means for modeling the transient high-pressure vaporization process of droplets of engine fuels. The calculated results characterize the vaporization processes of representative engine fuels, in this case diesel and gasoline droplets, and indicate the important influence of fuel volatility. The comparison with the results of single-component droplets also emphasizes the importance of considering multicomponent effects in practical applications, especially for high-pressure environments and volatile fuels. These results also demonstrate that the present model can be the basis for further study of the vaporization and combustion behavior of commercial petroleum fuels.

Acknowledgements

This work was supported by the DOE/Sandia National Laboratories and Caterpillar, Inc.

References

[1] Givler, S. D., and Abraham, J., 1996, ''Supercritical Droplet Vaporization and Combustion Studies,'' Prog. Energy Combust. Sci., **22**, pp. 1–28.
[2] Stengele, J., Bauer, H. J., and Wittig, S., 1996, ''Numerical Study of Bicomponent Droplet Vaporization in a High-Pressure Environment,'' ASME Paper No. 96-GT-442.
[3] Curtis, E. W., Uludogan, A., and Reitz, R. D., 1995, ''A New High Pressure Droplet Vaporization Model for Diesel Engine Modeling,'' SAE Paper No. 952431.
[4] Aggarwal, S. K., Shu, Z., Mongia, H., and Hura, H. S., 1998, ''Multicomponent and Single-Component Fuel Droplet Evaporation Under High-Pressure Conditions,'' AIAA Paper No. 98-3833.
[5] Zhu, G.-S., and Aggarwal, S. K., 1999, ''Fuel Droplet Vaporization in a Supercritical Environment,'' ASME Paper No. 99-GT-301.
[6] Zhu, G.-S., and Aggarwal, S. K., 2000, ''Droplet Supercritical Vaporization With Emphasis on Equation of State,'' Int. J. Heat Mass Transf., **43**, pp. 1157–1171.
[7] Abraham, J., and Givler, S. D., 1999, ''Conditions in Which Vaporizing Fuel Drops Reach the Critical State in a Diesel Engine,'' SAE Paper No. 1999-01-0511.
[8] Lazar, R. S., and Faeth, G. M., 1971, ''Bipropellant Droplet Combustion in the Vicinity of the Critical Point,'' *Proceedings 13th Symposium on Combustion*, Combustion Institute, Pittsburgh, PA, pp. 801–811.
[9] Hsieh, K. C., Shuen, J. S., and Yang, V., 1991, ''Droplet Vaporization in High-Pressure Environments I: Near Critical Conditions,'' Combust. Sci. Technol., **76**, pp. 111–132.
[10] Kuo, K. K., 1986, *Principles of Combustion*, John Wiley and Sons, New York.
[11] Turns, S. R., 1996, *An Introduction to Combustion*, McGraw-Hill, New York.
[12] Robert, E. C., John, E. D., Robert, M. G., and Dan, T. D., 1998, ''The Influence of Fuel Volatility on the Liquid-Phase Fuel Penetration in a Heavy-Duty DI Diesel Engine,'' SAE Paper No. 980510.
[13] Shayler, P. J., Colechin, M. J., and Scarisbrick, 1996, A., ''Fuel Film Evaporation and Heat Transfer in the Intake Port of an SI Engine,'' SAE Paper No. 961120.
[14] Tamim, J., and Hallett, W. L. H., 1995, ''Continuous Thermodynamics Model for Multi-Component Vaporization,'' Chem. Eng. Sci., **50**(18), pp. 2933–2942.
[15] Lippert, A. M., and Reitz, R. D., 1997, ''Modeling of Multi-Component Fuels Using Continuous Distributions With Application to Droplet Evaporation and Sprays,'' pp. 131–145, SAE Paper No. 972882.
[16] Lippert, A. M., 1999, ''Modeling of Multi-Component Fuels With Applications to Sprays and Simulation of Diesel Engine Cold Start,'' Ph. D. dissertation, University of Wisconsin, Madison, WI.
[17] Gal-Or, B., Cullinan, JR., H. T., and Galli, R., 1975, ''New Thermodynamic-Transport Theory for Systems With Continuous Component Density Distributions, Chem. Eng. Sci., **30**, pp. 1085–1092.
[18] Zhu, G.-S., and Reitz, R. D., 2001, ''A Model for High-Pressure Vaporization of Droplets of Complex Liquid Mixtures Using Continuous Thermodynamics,'' Int. J. Heat Mass Transf., to be published.
[19] Chou, G. F., and Prausnitz, J. M., 1986, ''Adiabatic Flash Calculations for Continuous or Semi-Continuous Mixtures Using an Equation of State, Fluid Phase Equilibria, **30**, pp. 75–82.
[20] Reid, R. C., Prausnitz, J. M., and Poling, B. E., 1987, *The Properties of Gases and Liquids*, McGraw-Hill, New York.
[21] Gradinger, T. B., and Boulouchos, K., 1998, ''A Zero-Dimensional Model for Spray Droplet Vaporization at High Pressures and Temperatures,'' Int. J. Heat Mass Transf., **41**, pp. 2947–2959.
[22] Chung, T. H., Ajlan, M., Lee, L. L., and Starling, K. E., 1988, ''Generalized Multi-parameter Correlation for Nonpolar and Polar Fluid Transport Properties,'' Ind. Eng. Chem. Res., **27**, pp. 671–679.
[23] Byung, I. L., and Michael, G. K., 1975, ''A Generalized Thermodynamic Correlation Based on Three-Parameter Corresponding States,'' J. AIChE, **21**, pp. 510–527.
[24] Knapp, H., Doring, R., Oellrich, L., Plocker, U., and Prausnitz, J. M., 1982, ''Vapor-Liquid Equilibria for Mixture of Low Boiling Substances,'' *Chemical Engineering Data, Series, VI*, DECHEMA, Frankfurt.
[25] Siebers, D. L., 1999, ''Scaling Liquid-Phase Fuel Penetration in Diesel Sprays Based on Mixing Limited Vaporization,'' SAE Paper No. 1999-01-0528.
[26] Higgins, B. S., Mueller, C. J., and Siebers, D. L., 1999, ''Measurements of Fuel Effects on Liquid-Phase Penetration in DI Sprays,'' SAE Paper No. 1999-01-0519.

G. M. Bianchi

P. Pelloni
Department of Mechanical Engineering,
University of Bologna, Italy

F. E. Corcione

L. Allocca
CNR-Istituto Motori,
Napoli, Italy

F. Luppino
VM-Motori,
Cento, Italy

Modeling Atomization of High-Pressure Diesel Sprays

This paper deals with a numerical and experimental characterization of a high-pressure diesel spray injected by a common-rail injection system. The experiments considered a free non-evaporating spray and they were performed in a vessel reproducing the practical density that characterizes a D.I. diesel engine at injection time. The fuel was supplied at high pressure by a common-rail injection system with a single hole tip. The computations have been carried out by using both the TAB model and a hybrid model that allows one to describe both liquid jet atomization and droplet breakup. In order to validate the breakup model, an extensive comparison between data and numerical predictions has been carried out in terms of spray penetration, Sauter mean diameter, near and far spray cone angles, and spray structure. [DOI: [10.1115/1.1361110]

Introduction

Over the past few years, D.I. diesel engine spray-combustion system technology has been dramatically improved in order to meet the more and more stringent pollutant emissions and fuel consumption standards. Diesel combustion is highly dependent on effective liquid fuel atomization, which controls the evaporation rate and the air-fuel mixing, and therefore emission levels and brake power of the engines ([1–4]). A deep knowledge of spray dynamics as well as of the interaction between liquid and gas phases is necessary in D.I. diesel engines design. Currently the experimental-CFD integrated approach represents a cost-efficient way of designing the spray-combustion systems ([1,5–8]). Optical diagnostic techniques provide the necessary insight into spray dynamics and the combustion process as well as data for the validation of CFD models ([1,2,9–13]). Numerical simulations could give detailed information on spray dynamics and guidelines to measurements, reducing the number of experimental test cases ([1,3,14,15]). Since spray modeling still remains a critical part of ICE simulations, especially when dealing with high-pressure high-dense sprays ([6,7,14]), this work is aimed at assessing the accuracy and the limits of spray models by using an integrated numerical and experimental approach. A free non-evaporating spray from a common-rail fuel injection system, typical of H.S.D.I. passenger car diesel engines, has been experimentally characterized at 1200 bar of rail pressure. Three different ambient densities, representative of engine conditions at injection, have been considered. Measurements were used in order to validate the high-pressure spray models. Simulations have been carried out by using three different spray models: (a) the TAB model ([16]), in order to make clear its inadequacy in modeling high-pressure sprays; (b) a hybrid model, which has been previously developed in [6] and applied with success in in-line pump mechanical injection system spray simulations. The hybrid model allows accounting for flow conditions inside the nozzle hole, like cavitation and liquid turbulence, that are found to be important at high injection pressures; (c) an updated version of the hybrid model. Modifications to the hybrid model have been deemed necessary in order to predict with a certain level of accuracy the structure of sprays from high-pressure common-rail injection systems. Detailed comparisons with experiments and discussions of model accuracy are presented. It will be demonstrated that flow conditions inside the nozzle as well as droplet detachment from the liquid jet surface have to be taken into account in spray modelling.

Contributed by the Internal Combustion Engine Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the ICE Division July 11, 2000; final revision received by the ASME Headquarters Dec. 7, 2000. Associate Editor: D. Assanis.

Experimental Setup and Test Conditions

The spray has been generated by an electronic controlled common-rail (CR) injection system supplied by Bosch [2]. Available software has allowed to define the pressure of the fuel in the rail and the duration and timing of the injection.

The spray has been injected by an axial single-hole injector, 0.18 mm in diameter, $L/D=5.56$, in an optically accessible vessel controlled in pressure with a constant gas temperature of 300 K. The use of an axial single-hole nozzle is an exemplification of the experimental scheme that enables us to orient the injector body in the most convenient way with respect to the stroboflash/strobodrum axes. It did not affect the behavior of the emerging fluid because the pressure gap through the nozzle was unvaried. A commercially available injector has typically five to six holes, 0.18 mm in diameter, distributed around the space. Finally, the interference of liquid coming from other holes could interfere with the examined one, producing a loss of accuracy.

Three different ambient pressure values have been considered: 2.0, 3.5, and 5.0 MPa.

The vessel was filled with nitrogen at a density varying from 1.17 Kg/m^3 to 58.48 Kg/m^3 at ambient temperature. The injection pressure was 120 MPa and the delivered quantity per stroke 11.825 10^{-3} g for a nominal injection duration of 1.0 ms.

A calibration fluid (ISO 4113) has been used to simulate the diesel fuel. The hydrodynamic properties at 15°C and atmospheric pressure of the ISO 4113 fluid are density $\rho_1=825\,\mathrm{kg/m^3}$, kinematic viscosity $\nu_1=2.5\times10^{-6}\,\mathrm{m^2/s}$, surface tension $\sigma_1=0.0284\,\mathrm{N/m}$. Fluid mass injected per single shot at $P_{inj}=120\,\mathrm{MPa}$ is 11.825×10^{-3} g.

An electronic signal synchronized the CR injection with the high speed photography apparatus (ImpulsPhysik GmbH).

A stroboflash lamp (20,000 fl/s) has frozen the emerging fuel from the nozzle on a Kodak 3200 ISO film rotating on a drum at 3600 rpm. The temporal resolution can be adjusted up to 50 μs. The prints of the processed film have been analyzed by Image ProPlus professional software for bidimensional image processing intensity of the transmitted light ([12,13]).

Dealing with the errors in measurements, it must be noted that the tip has been defined as the axial distance between the tip of the nozzle and the envelope of the spray. The resolution was $+/-1$ pixel and, taking into account the magnification ratio of the optical apparatus and the smoothing of the boundary of the spray, it did not exceed $+/-0.5$ mm.

The error on the spray-cone angle was practically negligible. In fact as spray completely developed ($t<500\ \mu$s), the straight lines, centered at the nozzle exit and developing along the boundary of the jet and compensating for the surface instability waves, were uniquely defined. In the image process the zoom function was largely used. An estimation of the spray-cone angle could be better than 0.2 deg.

Numerical Methodology

CFD Code and Grid. The computations were carried out by using the CFD KIVA-3 code ([1]) by modeling the experimental bomb with a grid size of $0.5\times0.5\times1$ mm^3 in the region close to the nozzle exit, as shown in Figure 1. The effect of grid resolution has also been investigated by considering coarser grids with a cell size typical of ICE calculations.

Fluid Dynamics Initial Conditions. Initial quiescent flow conditions have been assumed. Turbulence effects have been modeled by using the standard k-ε model, modified to account for compressibility. The initial kinetic energy k of the gas-phase was assumed to be uniform throughout the computational domain and equal to 60 cm^2/s^2. The initial dissipation rate ε was initialized by using a uniform turbulence length scale equal to 0.01 cm.

Spray Sub-Models. Droplet collision and coalescence are calculated according to the approach used in the KIVA-3 code ([17]). In modeling the dispersion of the particles by turbulence, the droplet fluctuation velocity was evaluated assuming an isotropic Gaussian distribution with a variance $\sigma^2=2/3k$, k being the turbulence kinetic energy of the gas phase. The drop drag coefficient is calculated according to ([18]). The total number of injected parcels is 20,000 for the TAB and the original hybrid models and 12,000 for the updated hybrid model because in this case new computational parcels are created during the simulation.

Injection Methodology. According to the method proposed in ([4]) (i.e., it is assumed that the dynamics and the breakup of a liquid column are indistinguishable from those of a train of droplets), the liquid injection is simulated by injecting blobs whose size is equivalent to the nozzle effective diameter (i.e., so-called ''blob injection'' methodology).

Nozzle and Injection Parameters. The global discharge coefficient c_D has been set equal to 0.65 from experiment. In order to assign a blob size and the maximum injection velocity, this value has been determined as the product of a contraction area coefficient c_a, assumed equal to 0.73 ([1]), and a flow rate reduction coefficient c_f determined equal to 0.9 from experiments of Chaves et al. [11]. Because of the lack of measurements, injection velocity has been evaluated by assuming that the maximum injection velocity is that given by Bernoulli's equation multiplied by the flow rate reduction coefficient c_f. The slope of the injection velocity curve during the needle opening and closing period has been derived by Tennison et al. [19] for the same injector and rail pressure as used in this research. Pressure fluctuations in the rail have been neglected.

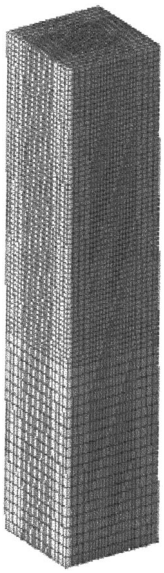


Fig. 1 Representation of the grid used in the calculations

Spray Model Description

TAB Model. The TAB model is a droplet secondary breakup model based on the so-called ''vibrational'' drop breakup mode. Droplet oscillation and distortion are modelled by using a simple forced harmonic oscillator, based on the analogy suggested by Taylor between an oscillating and distorting droplet and a spring mass system ([16]). The aerodynamics force is analogous to the external force F, the surface tension is analogous to the spring restoring force, while the damping force is related to the liquid viscosity. The governing equation of such a system is the

$$m\cdot\ddot{x}=F-k\cdot x-\zeta\cdot\dot{x}, \tag{1}$$

where x is the displacement of the droplet equator from its equilibrium position, m the mass of the droplet, k the spring-equivalent elastic constant, and ζ the system dumping constant. The characteristic oscillation frequency ω of the droplet that controls the breakup time is given by

$$\omega^2=C_k\cdot\frac{\sigma_l}{\rho_l\cdot r^3}, \tag{2}$$

where ρ_l and σ_l are the density and the surface tension of the liquid, respectively. According to O'Rourke and Amsden [16], the value of the constant C_k is 8. Further details about this model can be found elsewhere ([6,16]). The exclusive use of the TAB model or its modified versions in high-pressure spray calculations is really not appropriate since the atomization of the injected ''blobs'' is in effect modelled by a secondary breakup model as TAB is. The application of the TAB model (or modified versions) to simulate also the atomization process is justified elsewhere ([14]) by the attempt to overcome the problem of the definition of a prescribed droplet size distribution for a secondary breakup calculation. In this work, the use of TAB in conjunction with the so-called ''blob'' injection methodology is presented just as an example of its inadequacy in high-pressure spray applications.

Hybrid Breakup Model. This model was presented in a previous work ([6]) and it was applied with success to an in-line pump injection system. It will be referred to as the *HM*-1 model. In the following, a brief description of the model will be given. Further details can be found elsewhere in [6]. The hybrid breakup model considers both the atomization of the liquid jet and the droplet breakup (secondary breakup).

Atomization Model. The atomization process is supposed to be the result of the simultaneous effects of liquid jet cavitation and turbulence and aerodynamic interaction with the surrounding gas.

Turbulence-induced atomization sub-model. To account for the effects of the liquid jet turbulence on atomization, a modified version of the model of Huh and Gosman [15] is used. The model postulates that the turbulence fluctuations in the jet are responsible for the initial perturbations on the jet surface. Then, these waves grow according to Kelvin-Helmholtz instabilities until they detach as atomized droplets. The time scale of atomization is assumed to be a linear function of two time scales, τ_T and $\tau_{w,T}$,

$$\tau_{A,T}=\tau_{\mathrm{spn},T}+\tau_{\mathrm{exp},T}=C_1\cdot\tau_T+C_4\cdot\tau_{w,T}, \tag{3}$$

where τ_T is the turbulent time scale for the spontaneous growth time $\tau_{\mathrm{spn},T}$ and $\tau_{w,T}$ is the wave growth time for the exponential growth time $\tau_{\mathrm{exp},T}$.

The time scale $\tau_{w,T}$ is derived from the Kelvin-Helmholtz instability theory on an infinite plane for an inviscid liquid. The length scale of the turbulence L_T is assumed to be the dominant length scale of the atomization process: hence, the atomization length scale $L_{A,T}$ and the wavelength of surface perturbations $L_{W,C}$ are expressed as functions of L_T:

$$L_{A,T}=C_2\cdot L_T,\quad L_w=C_3\cdot L_T. \tag{4}$$

Assuming that half a surface wave is detached as a secondary drop from the jet, $L_{A,T}$ and $L_{W,T}$ are related by

$$L_{A,T}=0.5L_{W.T}. \tag{5}$$

Introducing Eq. (5) into Eq. (4), the constant C_3 can be expressed as a function of C_2 (i.e., $C_3=2\cdot C_2$). Due to the lack of turbulence measurements for fuel injectors, a simple force balance, based on the pressure drop along the nozzle downstream length, allows one to estimate the liquid jet average turbulent kinetic energy K_{avg} and its dissipation rate $\varepsilon_{\mathrm{avg}}$ at the nozzle exit: The liquid jet turbulent length and time scales are expressed as

$$L_T=c_\mu\frac{K_{\mathrm{avg}}^{3/2}}{\varepsilon_{\mathrm{avg}}}, \tag{6}$$

$$\tau_T=c_\mu\frac{K_{\mathrm{avg}}}{\varepsilon_{\mathrm{avg}}}. \tag{7}$$

The constant c_μ is equal to 0.09. The turbulence time and length scales of the parent drops (i.e., liquid jet) during their downstream motion are assumed to be constant contrary to the original model by Huh and Gosman [15], where they are assumed to vary with time. This assumption has been supported by calculations which show a very short parent-drop lifetime due to the fast atomization process typical of the high-pressure sprays which makes negligible the variations of L_T and τ_T with respect to their initial values computed at the nozzle exit (i.e., Eqs. (6) and (7)). The model constants have been set according to Huh and Gosman [15]: $C_1=1.0$, $C_2=2.0$, $C_4=1.5$.

Cavitation-induced atomization sub-model. Experiments reveal that cavitation or more frequently super-cavitation takes place at the nozzle exit under high injection pressure operating conditions ([1,9,10,20]). The sub-model tries to give an estimation of the cavitation-induced atomization characteristic time as the sum of two terms:

$$\tau_{A,C}=\tau_{\mathrm{spn},C}+\tau_{\mathrm{exp},C}=\tau_{\mathrm{cav}}+C_5\cdot\tau_{w,C} \tag{8}$$

where τ_{cav} is the cavitation time scale for the spontaneous growth time $\tau_{\mathrm{spn},C}$ and $\tau_{W,C}$ is the wave growth time for the exponential growth time $\tau_{\mathrm{exp},C}$. With respect to the original approach presented by Arcoumanis et al. [1] and according to Bode et al. [20], Eq. (8) allows one to account for the time required for the perturbations to become unstable before determining the atomization outcome.

In Eq. (8), $\tau_{\mathrm{spn},C}$ is calculated according to Arcoumanis et al. [1] and represents the time needed for the spontaneous growth of the perturbations induced by bubble collapsing or bursting. As in Arcoumanis's model, the characteristic time scale of the cavitation is assumed to be the smaller of the collapse time and the burst time:

$$\tau_{\mathrm{cav}}=\min(\tau_{\mathrm{coll}};\tau_{\mathrm{burst}}), \tag{9}$$

where the collapse time τ_{coll} of a bubble with radius R_{cav} is evaluated from the Rayleigh theory ([21])

$$\tau_c=0.9145\cdot R_{\mathrm{cav}}\sqrt{\frac{\rho_L}{p_v}}. \tag{10}$$

The burst time, calculated taking into account the time required for the cavities to reach the jet surface, is defined as

$$\tau_{\mathrm{burst}}=\frac{r_{\mathrm{hole}}-R_{\mathrm{cav}}}{u_L'}, \tag{11}$$

where u_L' is the velocity jet fluctuation ($u_L'=\sqrt{2K_{\mathrm{avg}}/3}$). R_{cav} is the radius of an equivalent bubble having the same area as all the cavitating bubbles, defined as

$$R_{\mathrm{cav}}=\sqrt{r_{\mathrm{hole}}^2-r_e^2}, \tag{12}$$

where r_e is the radius of the effective flow area at the nozzle exit and r_{hole} is the geometrical nozzle radius. The exponential growth time scale can be formally derived assuming that perturbations undergo Kelvin-Helmholtz instabilities; hence, $\tau_{W,C}$ can be expressed as

$$\tau_{w,C}=\frac{1}{\left[\dfrac{\rho_l\cdot\rho_g}{(\rho_l+\rho_g)^2}\left(\dfrac{u}{L_{w,C}}\right)^2-\dfrac{\sigma_l}{(\rho_l+\rho_g)\cdot L_{W,C}^3}\right]^{0.5}}, \tag{13}$$

where ρ_l and ρ_g are the liquid and gas density, respectively, while σ_l is the surface tension of the liquid fuel. The length scale of the cavitation-induced atomization $L_{A,C}$ is estimated according to Arcoumanis et al. [1]

$$L_{A,C}=2\pi(r_{\mathrm{hole}}-R_{\mathrm{cav}}). \tag{14}$$

Atomization characteristic time. In order to simulate the disintegration of the liquid jet, it is necessary to determine which is the driving process. Based on the experimental results of Fath et al. [9], it is assumed that, within a distance D_{ex} of 400 μm from the nozzle exit, cavitation and turbulence are competing and atomization is driven by the process that presents the lower atomization characteristic time scale. Beyond that distance, the atomization is supposed to be controlled by turbulence only. Hence, the atomization time scale is given by

$$\tau_A=\begin{cases}\min(\tau_{A,T};\tau_{A,C}), & D_{\mathrm{ex}}\leqslant 400\ \mu\mathrm{m}\\ \tau_{A,T}, & D_{\mathrm{ex}}>400\ \mu\mathrm{m}\end{cases}. \tag{15}$$

The choice of 400 μm as a criterion can be crude if one considers that cavitation is influenced by operating conditions and nozzle geometry. Justifications for the use of that value lie in the fact that the experiment of Fath et al. [9] has been carried out under conditions typical of a part load injection of practical diesel engines.

Owing to the jet breakup, the diameter of the blob injected is assumed to decrease *continuously* with time according to the following law:

$$\frac{dr}{dt}=-\frac{L_A}{\tau_A}, \tag{16}$$

where the atomization length scale is equal to $L_{A,T}$ or $L_{A,C}$ depending on which process is controlling the atomization. To conserve mass, the number of drops N in the computational parcel of radius r after the breakup event is adjusted with time according to $N_0r_0^3=Nr^3$ (N_0 is the number of parent drops of radius r_0 in the computational parcel). The spray cone angle ϑ is calculated as

$$\tan\frac{\vartheta}{2}=\frac{L_A/\tau_A}{U}, \tag{17}$$

where U is the relative velocity between liquid jet and surrounding gas.

Secondary Breakup Model. According to Habchi et al. [7] the blob Weber number We ($We=\rho_G dU^2/\sigma_L$, ρ_G being the gas density, d the particle diameter, U the relative velocity, and σ_L the surface tension) is taken as decisional criterion to distinguish between atomization and secondary breakup regimes. In this model,

it is assumed that all parent blobs having a Weber number $We \geq 1000$ undergo the atomization process, as suggested by Habchi et al. [7] while the secondary breakup model processes all droplets having a Weber number lower than the limit value of 1000. It must be pointed out that in reality the Weber number is not representative of distinguishing between liquid jet breakup and drop breakup. The Weber number as a criterion is used in calculations only because the liquid column emerging from the nozzle hole is simulated by a train of parent or ''blob'' droplets. The secondary breakup is simulated by the TAB model with constant C_k set equal to 1.0, according to Beatrice et al. [22] and Bianchi and Pelloni [6].

Updated Hybrid Model. In the very early period of injection, when the needle is opening, the blob Weber number could be lower than 1000; hence, *HM*-1 may treat the very first injected blobs by using the TAB model.

In order to overcome this weakness, an update to the previous hybrid models was introduced and all injected blobs were treated by the atomization sub-model previously presented. In a manner similar to that proposed by Reitz [23], atomization is simulated postulating that new drops with radius r_c are created from a parent (or blob) droplet with radius r according to

$$r_c = C_o L_A, \tag{18}$$

where C_o is a constant taken equal to 0.4. The reduction in size of the parent drop is assumed to follow Eq. (16). During the primary breakup events driven by Kelvin-Helmholtz instabilities, shedding-off of small droplets from the injected parent blobs occurs and new computational parcels are formed. In this work, the model by Su et al. [24] has been used in order to account for the mass of the new small droplets using the following equation:

$$m_{\text{shed}} = \frac{4}{3}\pi\rho_1 \sum_{i=0}^{j} n_i (r_i^3 - r_{i+1}^3), \tag{19}$$

where m_{shed} is the mass shed from the parent parcel, ρ_l is the density of the liquid fuel, n_i is the number of droplets in the parent parcel at time step i, and r is the radius of the droplets in the parent parcel in the subscripted time step. The summation starts at $i=0$, which is the time when the droplets begin the breakup process and end at term j, when enough mass exists to create a new parcel. The optimum threshold value of mass that allows new parcel creation has been set equal to 20 percent of the injected parcel mass based on those used by Habchi et al. [7] in their model. This particular value of the percentage has been found to give better agreement in terms of predicted SMD and spray structure with respect to measurements. The breakup of child droplets formed after the shedding process is evaluated by using the TAB model with constant C_k set equal to 1.0, as in the original hybrid model. In the following, the updated hybrid model will be referred to as *HM*-2.

Numerical Results

In order to test the spray models, the three injection conditions experimentally investigated have been simulated. Hence, for a fixed rail pressure of 120 MPa and an ambient temperature of 300 K, three different ambient pressures have been considered: 2.0 (Case A), 3.5 (Case B), and 5.0 MPa (Case C).

Model Performance. In order to compare the model behavior, Case A has been simulated. Then the performance of the updated hybrid model has been investigated also in Cases B and C.

Figure 2 shows the tip penetration versus time calculated for Case A by using the TAB model, the hybrid model (*HM*-1), and the updated hybrid model (*HM*-2). Figures 3 and 4 compare the predictions of the three models in terms of overall spray SMD and far spray cone angle, respectively. Figures 2 and 4 present a comparison with experimental data while no measurements of the overall spray SMD were available. Figure 9 compares the experimental and calculated spray structures at different instants after the start of injection (A.S.O.I).

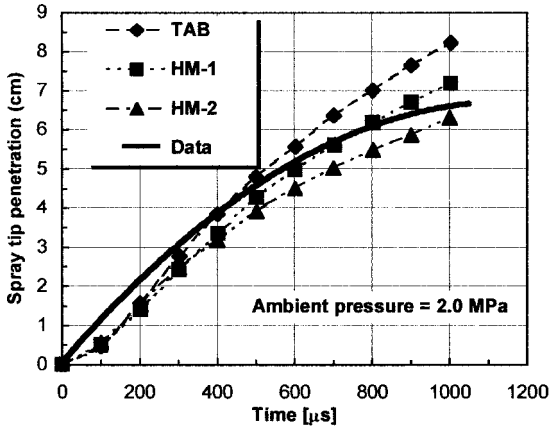


Fig. 2 Comparison between numerical and measured tip penetration-Case A

In the very early period from A.S.O.I., all three models underpredict the spray tip penetration, giving about the same results. Reasons for this discrepancy can be attributed to the sub-model used for evaluating the injection rate and to experimental errors. The injection sub-model seems to provide an underestimation of

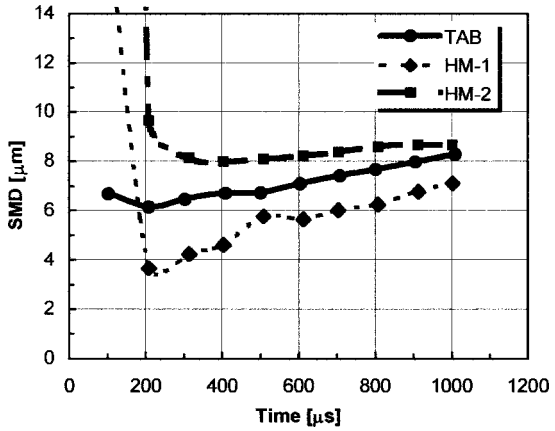


Fig. 3 Comparison between numerical predictions of overall spray SMD-Case A

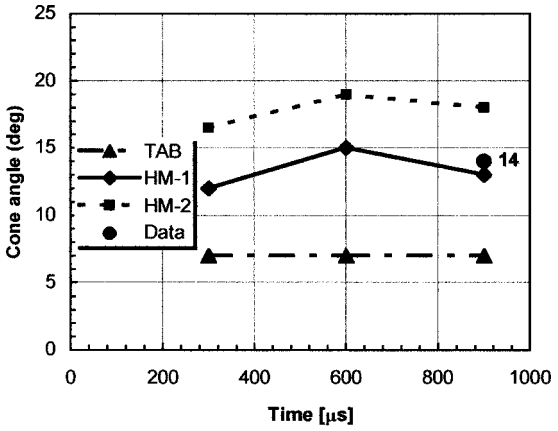


Fig. 4 Comparison between experimental and numerical predictions of far cone angles-Case A

the flow rate in the needle opening period due to uncertainties in determining the flow discharge coefficient in this highly unsteady phase. When the experiments are concerned, the temporal resolution of the pictures was 50 μs, which was also the uncertainties of the first frame in the spray evolution sequence.

Differences among model predictions occur from 300 μs A.S.O.I. The TAB model gives an overprediction of the spray tip penetration, leading to a difference of about 1.9 cm at 1000 μs A.S.O.I. The original and the updated hybrid models are much more accurate with respect to experiments: the little difference with measurements can be explained by experimental uncertainties. The latter are found to be significant in very early period from the start of injection (i.e., up to 200 μs). The original hybrid model provides more accurate predictions of the tip penetration than the updated version, as it will be confirmed later when Cases B and C will be discussed. A detailed investigation of the spray structure will show that this better agreement does not prove a superior capability of the original hybrid model because of inaccuracy in the spray structure prediction. Analysis of the spray tip penetration is not sufficient to validate the performance of a spray model: overall SMD (Fig. 3) and main cone angle (Fig. 4) must be checked too. Analysis of Fig. 3 confirms previous findings that the TAB model predicts a too short drop lifetime and therefore a too small spray SMD from the very early instants A.S.O.I.. As one can note in Fig. 3, at 100 μs after the start of injection, the TAB model gives a spray SMD of about 7 μm. If the influence of flow conditions on atomization rate are accounted for, as in *HM*-1 and *HM*-2, the spray SMD decreases more gradually, reaching in both cases the minimum value at about 200 μs A.S.O.I. Then, all models predict an increase of SMD due to coalescence processes, typical of high dense sprays. The use of particle shedding (*HM*-2) determines a lower atomization rate and therefore *HM*-2 predicts a minimum droplet size about twice as large as *HM*-1 at 200 μs A.S.O.I.. Analysis of Fig. 3 shows that the predictions of overall SMD evolution in time are consistent with SMD measurements of common rail sprays injected at the same conditions as those considered in the present work (see Tennison et al. [1] for a more detailed comparison). The *HM*-1 model gives better agreement with respect to the evolution of the far cone angle (Fig. 4), while the updated version of the hybrid models gives a slight overprediction. This behavior of *HM*-2 is typical of this case characterized by an ambient pressure of 2.0 MPa. As one can see in the following, the same tuning of the *HM*-2 model constants gives predictions of the far spray cone angle in very good agreement with experiments. By contrast, the TAB model predicts a very narrow far cone angle of 7 Deg that is far away from measurements.

The previous results are confirmed also in Cases B and C, characterized by an ambient pressure of 3.5 and 5.0 MPa, as shown in Fig. 5 and Fig. 6, respectively, where the measured and numerical evolutions of the spray tip penetration are compared for Cases B and C. Figure 7 shows a comparison between the measured and the predicted far spray cone angle by using the updated hybrid model. The agreement in terms of spray tip penetration still remains better if hybrid models are used, confirming the poor accuracy of the TAB model. Dealing with the far cone angle, it is seen that the updated hybrid model gives predictions in fairly good agreement with measurements in both cases, as shown in Fig. 7.

Figure 8 compares the spray structure predicted by the three models at 300, 600, and 900 μs A.S.O.I.. The spray images refer to the distribution of the droplet density expressed as grams of liquid per cm^3 of the gas phase. Figure 9 directly compares the calculated and the experimental spray shapes at 300, 600, and 900 μs A.S.O.I. for the *HM*-1 and *HM*-2 models. Analysis of Fig. 9 gives a useful insight into the model performance showing details of the spray structure that explain the reasons why the updated hybrid model seems to be more suited in high-pressure spray modelling than the original hybrid model and, especially, than the TAB model.

The previous analysis of the TAB predictions of spray penetration, SMD, and far cone angle have demonstrated significant inaccuracies. Analysis of Figs. 8 and 9 (the latter must be seen relative to the experimental part only) provides further doubts on the TAB model since it predicts a very narrow near cone angle of

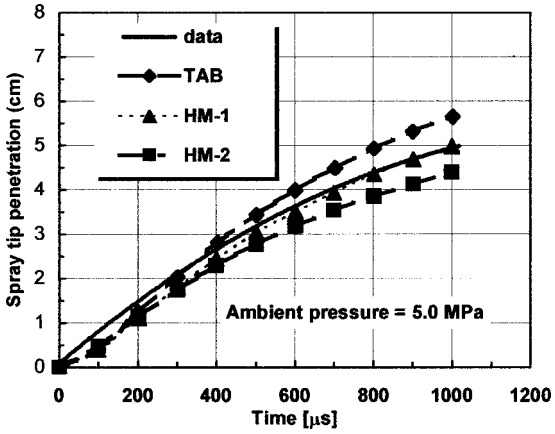


Fig. 6 Comparison between numerical and measured tip penetration-Case C

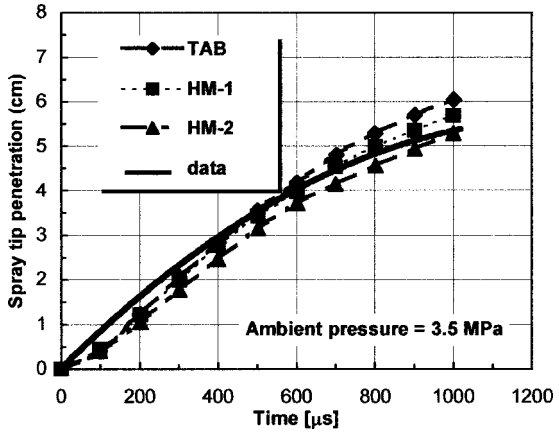


Fig. 5 Comparison between numerical and measured tip penetration-Case B

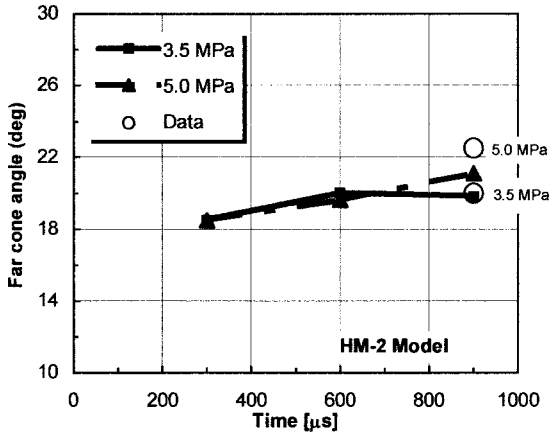


Fig. 7 Comparison between experimental and numerical predictions of far cone angle at different ambient pressures-*HM*-2 model

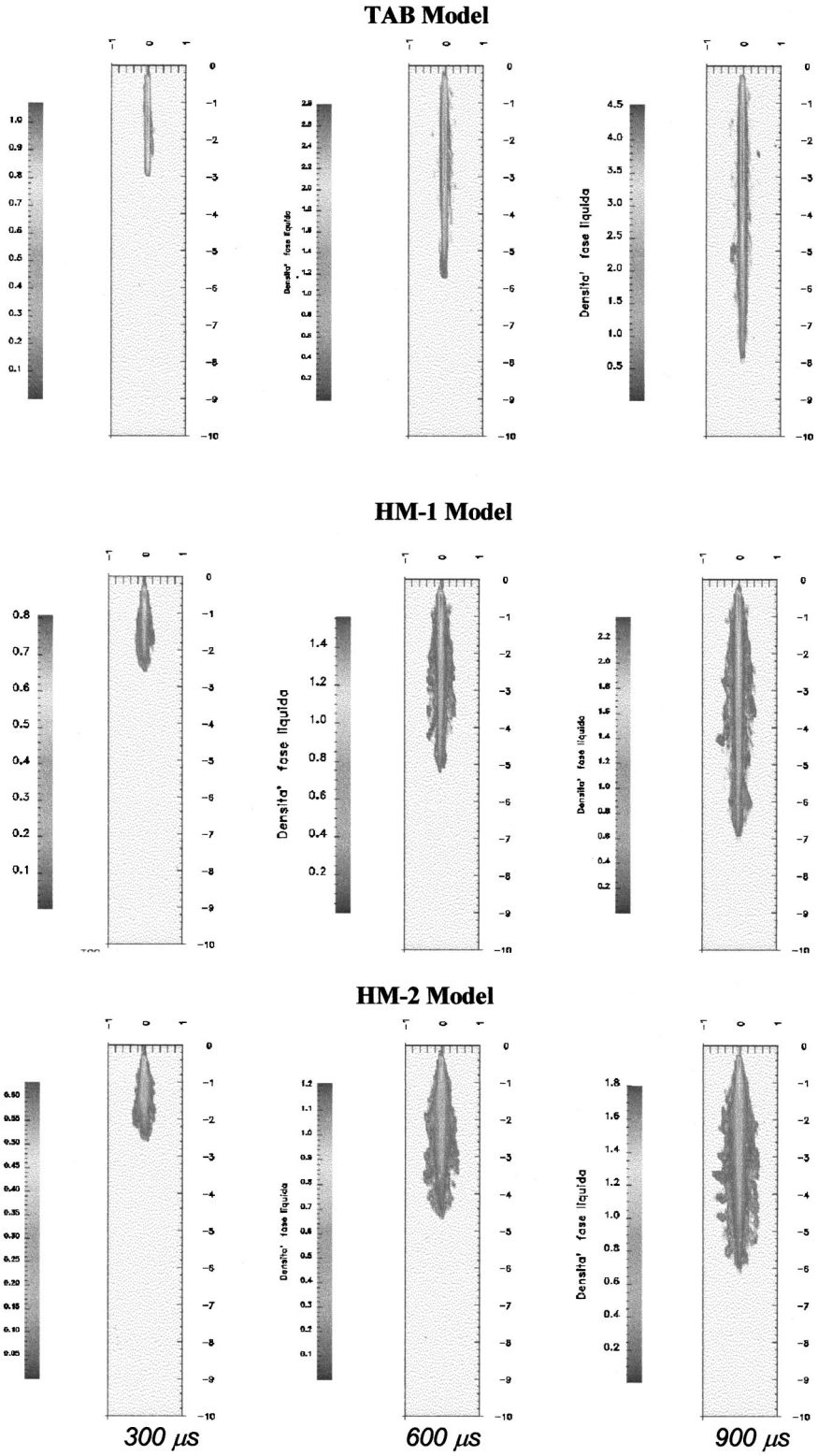


Fig. 8 Comparison of the breakup model influence on the spray density predictions. Plane passing through the spray axis.

about 10 Deg, three times lower than expected, as indicated by measurements of Tennison et al. [19] (i.e., about 30 Deg) for spray injection conditions like those examined in this analysis. Therefore, the TAB model is totally inaccurate in high-pressure spray modelling because of lack of physical ground. In fact, the TAB basic theory does not account for any of the main processes that drive liquid jet atomization (i.e., liquid jet cavitation and turbulence and aerodynamics interaction with surrounding gas). Beatrice et al. [22] have detected inaccuracies also in secondary breakup modeling suggesting an improved setting of the constant C_k.

This setting has been confirmed also by Bianchi and Pelloni [6],

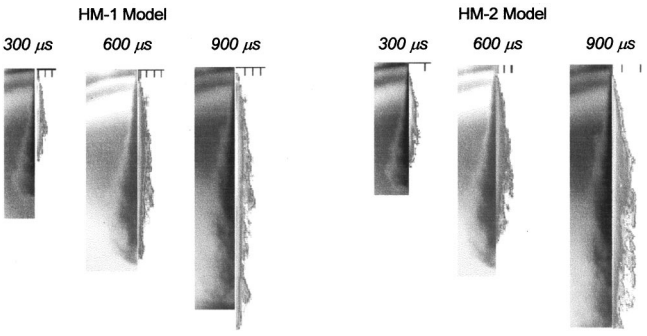


Fig. 9 Comparison between numerical (on the right) and experimental (on the left) spray structure

who demonstrated that the original value of C_k leads to an evaluation of the droplet breakup time that is almost an order of magnitude lower than the experimental one.

By contrast, when the influence of flow conditions inside the nozzle is taken into account, as in hybrid models, the shape of the spray close to the nozzle exit is well captured. As one can note in Figs. 8 and 9, *HM*-1 represents a great improvement with respect to the TAB model, while *HM*-2 provides even a better agreement predicting a near cone angle of about 30 Deg very close to the value reported in the literature by Tennison et al. [19]. The droplet shedding—off procedure used to simulate drop detachment from the liquid jet surface determines more accuracy in modeling the spray structure especially after 300 μs A.S.O.I. Comparing the experimental and the calculated structures of the spray at 600 μs and 900 μs A.S.O.I.. one can note that the application of *HM*-1 to common-rail sprays reveals inaccuracies in capturing both the inner and the outer region details. In particular, *HM*-1 predicts a narrow spray with a dense core extending from the nozzle exit to a distance of 2 cm from the injector. The presence of droplets in the periphery is coarse. Hence, the spray is less dispersed and droplet momentum is conserved along the spray axis determining good agreement in spray tip penetration predictions. On the contrary, significant improvements are achieved by using the *HM*-2 model. In particular, by using the droplet shedding procedure, a very dense region is predicted only close to the nozzle exit and the experimental characteristic conical shape in the inner region of the actual spray is captured. Moreover, since the droplet dispersion in the periphery is modelled with more accuracy, the spray shape results in good agreement with that given by experimental images. It must be pointed out that the *HM*-2 model slightly under-predicts the spray tip penetration with a maximum error of 9 percent.

Trying to draw some conclusions, calculation results and comparisons with experiments indicate that the influence of nozzle flow conditions must be taken into account in modelling high-pressure diesel sprays. The TAB model is therefore far from being accurate in such calculations. With reference to hybrid models, the update proposed to the hybrid model allows improving drastically the accuracy of computations in terms of spray structure. This is extremely important in spray combustion calculations because spray dynamics determines the air-fuel distribution and therefore the ignition point and the following combustion development.

Effect of Grid Resolution

Even if the results of previous calculations are in good agreement with experimental data, it is not possible to apply spray models without considering the effect of the grid resolution. This is even a more critical issue for high-pressure sprays since the numerical procedure usually adopted in spray computations assumes that the volume occupied by liquid is small with respect to the volume occupied by the gas phase. This could not be verified in high dense sprays, especially near the orifice. Following a common methodology, also in this research spray models were tested on a fine grid resolution of $0.5\times0.5\times1$ mm^3 in the *x*, *y*, and *z* directions, respectively, where *z* is coincident with spray axis. The finer resolution in the *x* and *y* directions is required because of the highest velocity gradients in the cross-spray sections. On the other hand, this resolution could provide numerical problems in high dense sprays ([25,26]). This grid is much finer than that usually adopted for I.C.E. computations, also when dealing with small-bore diesel engines. In fact, practical ''engine'' grids have cells of about of 1 mm^3. This resolution is not adequate to solve accurately the spray dynamics as shown by Abraham [25]. On the other hand, currently it is impossible to make use of relatively fine grids ($0.5\times0.5\times1$ mm^3) because of the limits in computational storage and run time.

In order to investigate the effects of the cell size on numerical results, three different mesh resolutions have been considered in simulating Case B: a *fine* grid ($0.5\times0.5\times1$ mm^3), which is the grid used for all previous calculations, a *medium* grid ($0.75\times0.75\times1$ mm^3), and a *coarse* grid ($1\times1\times1$ mm^3). The updated hybrid model has been adopted for these calculations. In Fig. 10, the influence of grid resolution on the tip penetration is presented. It is seen that the coarse grid predicts a lower penetration while the fine grid gives the best agreement with experimental data. The effect of grid resolution on SMD is shown in Fig. 11. Despite the fact that the values of SMD are reasonable for all the grids considered, it is possible to note that different grid resolutions determine different droplet size predictions. This is mainly due to the differences in the turbulence levels as one can note in Fig. 12, where the turbulence kinetic energy is compared for the fine and coarse grids. Using the coarse grid determines inaccuracies in

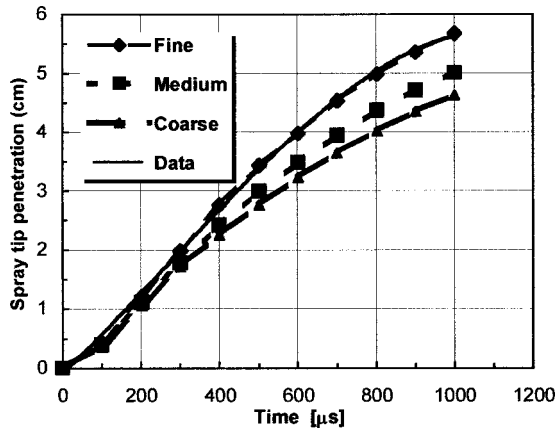


Fig. 10 Influence of grid resolution on tip penetration

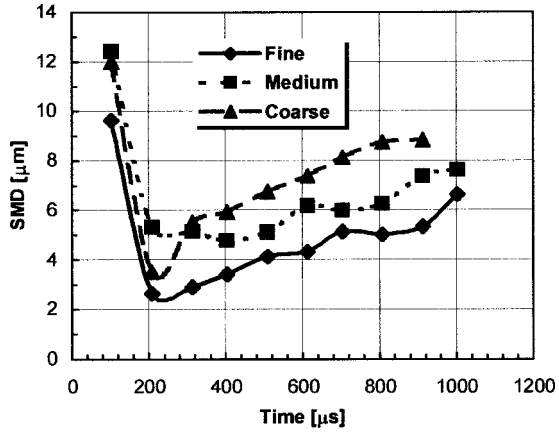
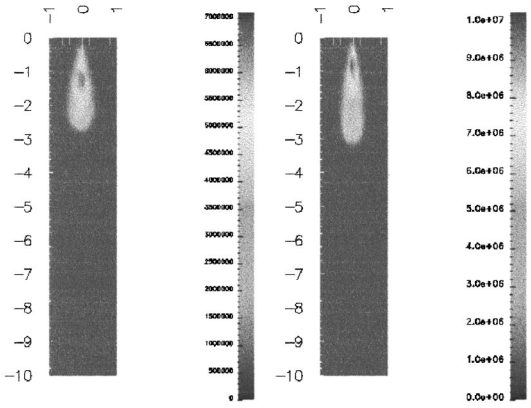
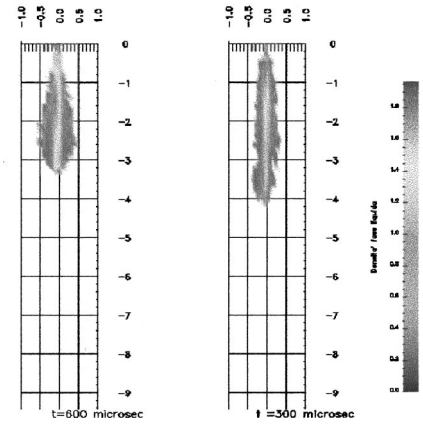


Fig. 11 Influence of grid resolution on SMD



Coarse grid *Fine grid*

Fig. 12 Effect of grid resolution on turbulence *kinetic energy*-600 μs A.S.O.I.



Coarse grid *Fine grid*

Fig. 13 Effect of grid resolution on spray *structure* 600 μs A.S.O.I.

solving the steep gradients in the gas phase induced by momentum transfer from the liquid phase that in turn affect atomization and breakup rate as well as droplet turbulent dispersion. As a result, the use of a coarse grid causes a radial dispersion of the spray with a reduction of the axial spray velocity, as one can see in Fig. 13, which compares the predictions of spray density on a plane passing trough the spray axis at 600 μs A.S.O.I. given by coarse and fine grids. As the grid becomes coarser, the spray presents a more dense core more downstream of the nozzle exit.

Summarizing the previous results, it is clear that the influence of the cell size is controversial because spray sub-models do not provide converged solutions as the grid is refined. Definitive conclusions are far from being drawn. Computation results seem to identify the problem of the convergence mainly in numerical aspects. (a) If the grid size gets closer to the nozzle hole diameter, the assumption that the volume occupied by liquid must be small with respect to the volume occupied by the gas phase is violated. Hence, the calculation of high-dense spray becomes numerically critical especially for the collision and coalescence sub-models. This weakness is visible in the results of this research, which show a dramatic sensitivity of the predicted spray SMD to the grid size (Fig. 11). (b) The second issue is that the Langragian-Eulerian approach used in spray-gas calculations overpredicts the momentum exchange between liquid and gas phase close to the nozzle orifice. This aspect can be eliminated by high grid refinements that allow a more accurate resolution of the high velocity gradients preventing the overestimation of momentum diffusion. Unfortunately, this procedure moves in the opposite direction than that required by the first issue previously addressed. What one can conclude from the above discussion is that the numerical solution of high-pressure spray equations needs to be addressed further. Currently, coarser grids cause an underprediction of spray penetration and an overprediction of the radial dispersion of the droplets, with consequent effects on the air/fuel distribution in vaporizing sprays and, in turn, on the combustion process. Hence, simulation and testing of spray models in a bomb with a relatively fine grid could not be enough to draw firm conclusions when spray models are applied in engine spray combustion calculations with a relatively coarse grid.

Summary and Conclusions

This research has been focused on an integrated numerical and experimental study of spray from a common-rail system for passenger cars. The aim was to develop a reliable model for the simulation of liquid jet atomization and droplet breakup of high-pressure diesel fuel spray.

The experiments considered a free non-evaporating spray and they were performed in a vessel reproducing the practical densities that characterize passenger car H.S.D.I. diesel engines at injection time.

The extensive comparison with experiments has shown that flow conditions inside the nozzle and simulation of droplet detachment from the liquid jet surface have to be modeled to achieve fairly accurate predictions of the main spray parameters.

Grid resolution strongly affects the numerical simulation when dealing with high-pressure spray because the spray models tested seem to not provide converged solutions. Aspects of the numerical solution of high-pressure spray equations need to be addressed in the future.

Acknowledgments

This work has been supported by VM Motori (Detroit Diesel Corporation), MURST, and CNR. The authors would like to acknowledge the contribution of Mr. A. Basile to the development of the spray models.

References

[1] Arcoumanis, C., Gavaises, M., and French, B., 1997, ''Effect of Fuel Injection on the Structure of Diesel Sprays,'' SAE Paper No. 970799.
[2] Boehner, W., and Hummel, K., 1997, ''Common Rail Injection System for Commercial Diesel Vehicles,'' SAE Paper No. 970345.
[3] Patterson, M. A., and Reitz, R. D., 1998, ''Modeling the Effects of Fuel Sprays Characteristic on Diesel Engine Combustion and Emission,'' SAE Paper No. 980131.
[4] Han, Z., Uludogan, A., Hampson, G. J., and Reitz, R. D., 1996, ''Mechanism and NOx Emissions Reduction Using Multiple Injection in a Diesel Engine,'' SAE Paper No. 9606336.
[5] Allocca, L., Belardini, P., Bertoli, C., Corcione, F., and De Angelis, F., 1992, ''Experimental and Numerical Analysis of a Diesel Spray,'' SAE Paper No. 920576.
[6] Bianchi, G. M., and Pelloni, P., 1999, ''Modeling the Diesel Fuel Spray Breakup by Using a Hybrid Model,'' SAE Paper No. 1999-01-0226.
[7] Habchi, C., Baritaud, T., et al., 1997, ''Modeling Atomization and Break Up in High Pressure Diesel Sprays,'' SAE Paper No. 970881.
[8] Reitz, R. D., 1987, ''Modeling Atomization Process in High-Pressure Vaporing Sprays,'' Atomization Spray Technol. **3**, pp. 309–337.
[9] Fath, A., Munch, K. U., and Leipertz, A., 1997, ''Spray Break-Up Process of Diesel Fuel Investigated Close to the Nozzle,'' Proceedings of ICLASS-'97, Aug. 18–22, Seoul, Korea.
[10] Faeth, G. M., 1990, ''Structure and Atomization Properties of Dense Turbulent Sprays,'' *Proc. 23rd Symp. (Int.) on Combustion*, The Combustion Institute, Pittsburgh, PA, pp. 1345–1352.
[11] Chaves, H., Knapp, M., Kubitzek, A., Obermeier, F., and Schneider, T., 1995, ''Experimental Study of Cavitation in the Nozzle Hole of Diesel Injectors Using Transparent Nozzles,'' SAE Paper No. 950290.
[12] Alfuso, S., Allocca, L., Di Stasio, S., and Corcione, F. E., 1999, ''Image Analysis of a Common Rail Injection Spray Evolving Into a Quiescent Ambient at Different Backpressures,'' Combustion Meeting 1999-XXII Event of the Italian Section of the Combustion Institute, Florence, 2–5 May.

[13] Corcione, F. E., Allocca, L., Vaglieco, B. M., and Valentino, G., 1998, ''Spray Formation in Diesel Combustion Systems,'' *IWASC 98*, Hiroshima, Japan, pp. 73–82.
[14] Tanner, F. X., and Weisser, G., 1998, ''Simulation of Liquid Jet Atomization of Fuel Sprays by Means of a Cascade Drop Breakup Model,'' SAE Paper No. 980808.
[15] Huh, K. Y., and Gosman, A. D., 1991, ''A Phenomenological Model of Diesel Spray Atomization,'' *Proceedings of the International Conference on Multiuphase Flows*, Sept. 24–27, Tsukuba, Japan.
[16] O'Rourke, P. J., and Amsden, A. A., 1987, ''The Tab Method for Numerical Calculation of Droplet Breakup,'' SAE Paper No. 870289.
[17] Amsden, A. A., O'Rourke, P. J., and Butler, T. D., 1993, ''Kiva-3: A KIVA Program with Bock-Structured Mesh for Complex Geometries,'' Los Alamos National Labs, LA-12503-MS.
[18] Liu, A. B., Mather, D., and Reitz, R. D., ''Modeling the Effects of Drop Drags and Breakup on Fuel Sprays,'' SAE Paper No. 930072.
[19] Tennison, P., Georjon, T. L., Farrell, P. V., and Reitz, R. D., 1998, ''An Experimental and Numerical Study of Sprays From Common Rail Injection System for Use in HSDI Diesel Engines,'' SAE Paper No. 980810.
[20] Bode, J., Chaves, H., Obermeier, F., and Schneider, T., 1991, ''Influence of Cavitation in Turbulent Nozzle Flow on Atomization and Spray Formation of a Liquid Jet,'' *Proceedings, Sprays and Aerosols 91*, Guildford, UK, pp. 107–112.
[21] Bayel, L., and Orzechowski, Z., 1993, *Liquid Atomization*, Taylor & Francis, London.
[22] Beatrice, C., Belardini, P., Bertoli, C., Cameretti, M. C., and Cirillo, N. C., 1995, ''Fuel Jets Model for Multidimensional Diesel Combustion Calculation: An Update,'' SAE Paper No. 950086.
[23] Reitz, R. D., 1996, ''Computer Modeling of Sprays,'' ERC, University of Wisconsin, Madison, WI.
[24] Su, T. F., Patterson, M. A., Reitz, R. D., and Farrell, P. V., 1996, ''Experimental and Numerical Studies of High Pressure Multiple Injection Sprays,'' SAE Paper No. 960861.
[25] Abraham, J., 1997, ''What is the Adequate Resolution in the Numerical Computation of Transient Jets?,'' SAE Paper No. 970051.
[26] Abraham, J., and Magi, V., 1999, ''A Virtual Liquid Source (VLS) Model for Vaporizing Diesel Sprays,'' SAE Paper No. 1999-01-0911.
[27] Allocca, L., Corcione, F. E., Fusco, A., Papetti, F., and Succi, S., 1994, ''Modeling of Diesel Spray Dynamics and Comparison With Experiments,'' SAE Paper No. 941895.
[28] Reitz, R. D., and Diwakar, R., 1997, ''Structure of High-Pressure Fuel Sprays,'' SAE Paper No. 870598.
[29] Gonzales, D., Borman, M. A., and Reitz, R. D., 1992, ''Modeling Diesel Engine Spray Vaporization and Combustion,'' SAE Paper No. 920579.

The Frequency Analysis of the Crankshaft's Speed Variation: A Reliable Tool for Diesel Engine Diagnosis

D. Taraza

N. A. Henein

Department of Mechanical Engineering,
Wayne State University,
5050 Anthony Wayne Drive,
Detroit, MI 48202

W. Bryzik

U.S. Army–TARDEC,
Mail Stop 121,
Warren, MI 48397-5000

The speed variation of the crankshaft may be easily and accurately measured using a shaft encoder and counting the pulses of the internal clock of the data acquisition system. If the crankshaft would be a rigid body, the variation of its angular speed could be directly correlated to the total gas-pressure torque. Actually, the variation of the crankshaft's speed has a complex nature being influenced by the torsional stiffness of the cranks, the mass moments of inertia of the reciprocating masses and the average speed and load of the engine. Analyzing only the lower harmonic orders of the speed variation spectrum can filter out the distortions produced by the dynamic response of the crankshaft. The information carried by these harmonic orders permits to establish correlations between measurements and the average gas pressure torque of the engine, and to detect malfunctions and identify faulty cylinders. [DOI: 10.1115/1.1359479]

Introduction

The methods devised to detect nonuniformities in the contributions of the cylinders to the total engine output, by analyzing the variation of the measured crankshaft's speed, may be classified in four groups: pattern recognition, statistical estimators, order domain methods, and a combination of these methods. A good overview of the current state of the art was presented by Williams [1].

Pattern recognition methods ([2,3]) require a large amount of experimental data obtained on the same engine operating with known faults at various speeds and loads. These reference patterns are used to form basic sets for comparison with measured speed waveforms. Statistical estimators methods are based on an engine roughness parameter ([4–7]), determined from the measured speed fluctuation. Sometimes, a model of the powertrain is considered and order domain analysis is used to identify the faulty cylinder(s) ([8–14]).

Order domain methods are suitable for steady-state operating conditions when all the factors that influence the crankshaft's speed variation may be considered as periodic functions of time (crank angle). In this case, the dynamic behavior of the crankshaft can be described, fairly accurately, by linear differential equations and, if an accurate dynamic model of the crankshaft is available, the calculations are in close agreement with the measurements ([12,14]). All these methods involve a large amount of calculation to detect and identify faulty cylinders.

The method developed in this paper is basically an order domain method, but it is based only on information supplied by the lowest harmonic orders of the measured speed. It could be also applied in cases when a dynamic model of the crankshaft is not available.

Crankshaft's Dynamics

A pertinent correlation between the angular motion of the crankshaft and the gas-pressure torque of each cylinder may be developed using the lumped-mass torsional dynamic model of the crankshaft. This model has been extensively used with very good results to predict and control torsional vibrations ([15,16]), to correlate the amplitude of the major harmonic orders of the crankshaft's speed variation with the average Indicated mean effective pressure (IMEP) of the engine ([14]) and to detect differences in the operation of individual cylinders, based on the measured crankshaft's speed variation ([8,9,11,13]).

The external torque acting on the crankshaft is mainly a contribution of the gas-pressure and reciprocating inertia torques, corresponding to each cylinder of the engine. Under steady-state operating conditions, the variation of the reciprocating inertia torque may be considered identical for all cylinders, and shifted in the crank-angle domain according to the firing order of the engine. The gas-pressure torque may differ from cylinder to cylinder, but, for steady-state operating conditions, the gas-pressure torque of a given cylinder undergoes little deviation from cycle to cycle. In this situation the torque corresponding to a cylinder is a periodic function of time (or crank angle) and may be expressed as a Fourier series. Usually 24 harmonic components are necessary to represent the gas-pressure torque for a four-stroke engine.

The resultant torque acting on the crankshaft is the sum of the individual torques contributed by each cylinder. This summation may be performed for each harmonic component separately, the resultant torque being obtained by adding the contribution of all harmonic components. The dynamic response of the crankshaft results from the superposition of the responses to each harmonic component of the resultant engine torque. While the higher harmonic components of the gas pressure torque may excite torsional vibrations of the shafting at different engine speeds, the frequencies of the lower harmonic components are much lower than the first natural frequency of the shafting for all operating speeds of the engine.

The typical phase angle diagrams of the lower harmonic orders of the resultant gas pressure torque for four-stroke, four- and six-cylinder engines with uniform firing intervals are shown in Figs. 1 and 2. The diagrams show that, if all cylinders operate identically, only the resultants of the major orders[1] will subsist in the frequency spectrum of the resultant torque ([12,16]). Actually, the operation of all cylinders cannot be absolutely identical and small contributions of nonmajor harmonic orders will be present in the spectrum of the resultant gas pressure torque.

Because the lower harmonic components of the resultant gas pressure torque do not excite torsional vibrations, the dynamic

Contributed by the Internal Combustion Engine Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the ICE Division Feb. 9, 2000; final revision received by the ASME Headquarters Dec. 6, 2000. Associate Editor: D. Assanis.

[1]The major orders, for a four-stroke engine, are multiples of half of the cylinder number.

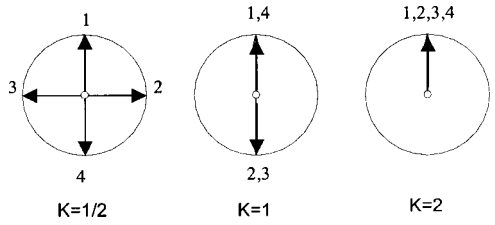


Fig. 1 Typical phase angle diagrams of a four-stroke, four-cylinder engine and the corresponding lowest harmonic order

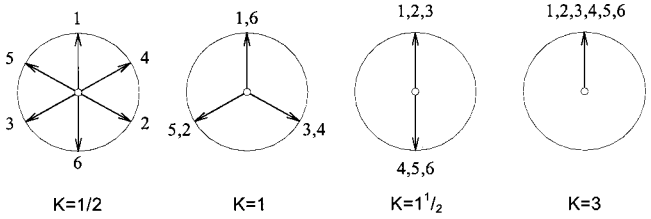


Fig. 2 Typical phase angle diagrams of a four-stroke, six-cylinder engine and the corresponding lowest harmonic order

response of the shafting to these harmonic orders should reflect their amplitudes, permitting to estimate the average gas pressure torque (or IMEP) and detect differences in the operation of the cylinders.

Experimental Investigation

Extensive experimentation was conducted on two four-stroke, direct injection diesel engines, a four-cylinder (DDC Series 50) and a six-cylinder (Cummins 6CTA 8.3).

The engines were run at constant speed and different loads, and several nonuniformities have been simulated. To simulate a faulty cylinder, the nut connecting the high-pressure fuel line to the corresponding element of the injection pump was slightly unscrewed and a leakage was introduced in the fuel supply of the cylinder. Controlling the tightness of the high-pressure fuel line, the amount of fuel injected into the cylinder may be reduced gradually from the rated value to zero, to simulate a wide range of nonuniformities up to a complete misfire.

The pressures were measured in all cylinders by piezoelectric pressure transducers (Kistler 6123) via charge amplifiers (Kistler 5010). The crankshaft's speed variation was determined using a high precision hollow shaft encoder (PEI-5VL670 HAZ 10) with 360 divisions on the optical disc. The hollow encoder was mounted on the flywheel and permitted the coupling of the engine to the dynamometer. An absolute pressure transducer (OMEGA PX176-025 A5V) was used to measure the crankcase pressure and all experimental data were sampled and recorded by a 24-channel data acquisition system (DSP Technology) having an internal clock with a 10 MHz frequency.

The IMEP of each cylinder was calculated from the pressure traces and the measured speed was subjected to a discrete Fourier transform (DFT) to determine the amplitudes and phases of its harmonic components.

Harmonic Structure of the Resultant Gas Pressure Torque

The resultant gas pressure torque was calculated from the pressure variation measured in each cylinder and then subjected to a DFT to obtain its spectrum. When all cylinders operate fairly uniform, the spectrum of the total gas-pressure torque contains mainly the major harmonic orders of the engine with insignificant contribution of other orders. This is true for both the four-cylinder engine (Fig. 3) and for the six-cylinder engine (Fig. 4).

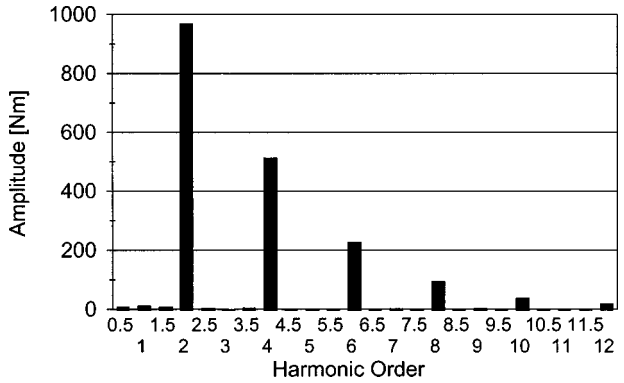


Fig. 3 Frequency spectrum of the resultant gas-pressure torque for the four-cylinder engine, idling at 720 rpm

Figure 3 shows the spectrum of the resultant gas pressure torque for the four-cylinder engine with all cylinders operating fairly uniform. The harmonic components of the resultant gas pressure torque are basically the major harmonic orders of the four cylinder engine ($k=2$, 4, 6, 8, 10, 12). Small differences in the cylinders' operation result in negligible contributions of the nonmajor harmonic orders. Compared to the amplitude of the second harmonic order ($k=2$) the maximum amplitude of a nonmajor order ($k=1$) is only about 1 percent.

The same situation is valid for the six-cylinder engine (Fig. 4) when the cylinders are contributing uniformly to the resultant gas pressure torque. On the contrary, when a large nonuniformity is introduced by suppressing the fuel injection into a cylinder, the lower nonmajor orders play a large role in the frequency spectrum of the resultant gas-pressure torque (Fig. 4). The analysis of Fig. 4 shows that, both for uniform and nonuniform contribution of the cylinders, the dominant component of the spectrum is the third harmonic order, which is the first major order of this engine. At the same time, the values of the amplitudes of the third harmonic order reflect the differences in the average engine torque for the two operating conditions.

The frequency spectra of the resultant gas pressure torque, for two different irregularities in the operation of the cylinders, but for the same value of the IMEP (or average torque) are presented in Fig. 5. Analysis of Fig. 5 shows that, in spite of the fact that the engine operates with completely different cylinder loadings, the amplitudes of the major harmonic orders of the average torque have almost the same value if the average torque (or IMEP) is the

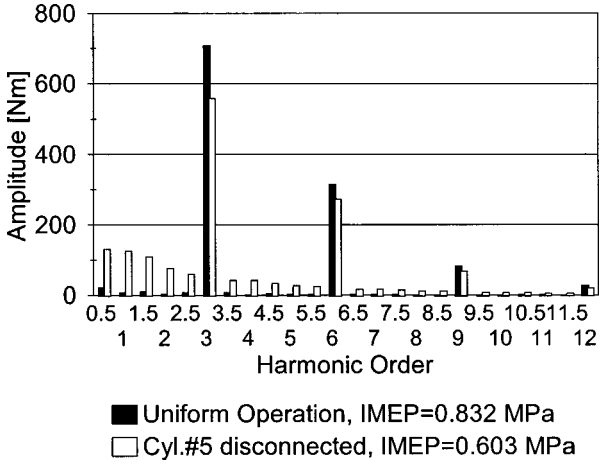


Fig. 4 Frequency spectra of the resultant gas-pressure torque. Six-cylinder engine, 1200 rpm.

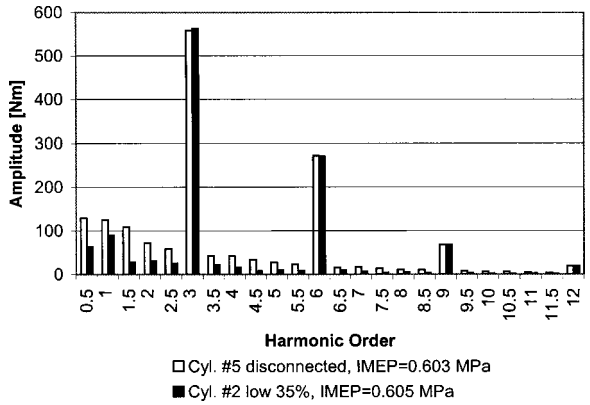


Fig. 5 Frequency spectra of the resultant gas-pressure torque. Six-cylinder engine, 1200 rpm.

same. Because in both cases shown in Fig. 5 there is a nonuniform contribution of the cylinders, the amplitudes of the lowest harmonic orders ($k=1/2$, $k=1$, and $k=1\frac{1}{2}$) are significant and reflect the degree of nonuniformity.

Estimation of the Average Gas-Pressure Torque From the Measured Crankshaft's Speed

If the lowest major order has a frequency that is far from the first natural frequency of the shafting, the amplitude of this order, in the crankshaft's speed spectrum, will be proportional to the amplitude of the corresponding order in the resultant gas-pressure torque spectrum. This dependency could be used to determine a quantitative relationship between the amplitude of the first major order of the engine, calculated by a DFT of the measured speed, and the average IMEP, or gas-pressure torque.

In Fig. 6, the frequency spectrum of the measured speed is presented for the two cases considered in Fig. 4. There is a very close relationship between the crankshaft's speed and gas-pressure spectra in both cases. The major difference occurs for the orders $k=7.5$ and $k=8$ that have a significant contribution in the speed spectrum, and are completely insignificant in the gas-pressure torque spectrum. This discrepancy is caused by the fact that the shafting is close to resonance with the eighth harmonic order when the engine is running at 1200 rpm. This situation points out the necessity to choose the lowest major order for the correlation to the average IMEP of the engine.

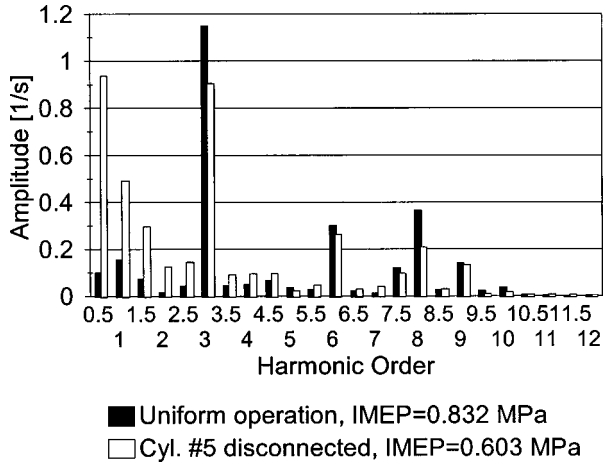


Fig. 6 Frequency spectra of the crankshaft's speed. Six-cylinder engine, 1200 rpm.

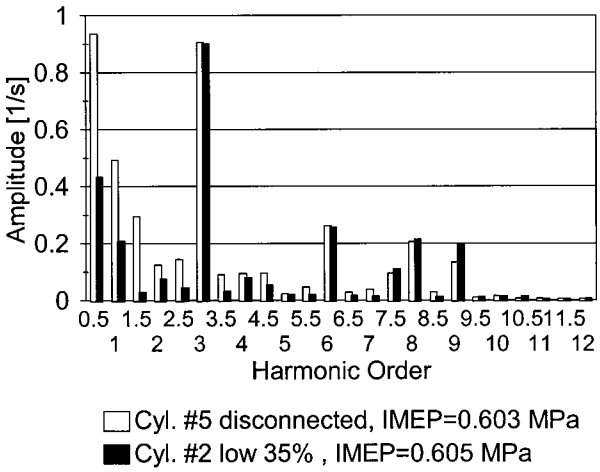


Fig. 7 Frequency spectra of the crankshaft's speed. Six-cylinder engine, 1200 rpm.

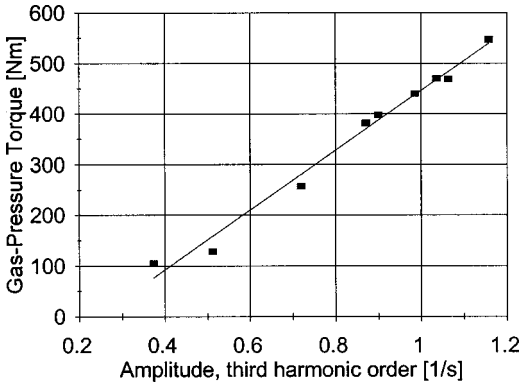


Fig. 8 Average gas-pressure torque as a function of the amplitude of the third harmonic order of the measured speed

An interesting fact is that, even if the contribution of the cylinders is very different, the amplitude of the lowest major harmonic order in the crankshaft's speed spectrum always reflects the average value of the IMEP. This property is further proved by comparing the amplitudes of the third order for two different operating conditions (Fig. 7) at the same speed (1200 rpm) and load (IMEP≅0.604 MPa).

Based on this property, the experimental six-cylinder engine was operated at 1200 rpm with different loads and, both with uniform and nonuniform contribution of the cylinders. The results are presented in Fig. 8.

Inspection of Fig. 8 shows that, for a given engine speed, the dependency of the average gas-pressure torque on the amplitude of the third harmonic order of the measured crankshaft's speed is quite linear. Similar linear dependencies may be obtained for other engine speeds to build a complete map for correlating the average gas-pressure torque (or IMEP) to the amplitude of the third harmonic order of the measured speed.

Direct Algorithm for Detecting NonUniform Cylinder Operation

When the cylinders are uniformly contributing to the total engine torque, the first three harmonic orders ($k=1/2$, $k=1$, and $k=1\frac{1}{2}$) play an insignificant role in the frequency spectrum of the total gas-pressure torque (see Figs. 3 and 4) and, consequently, appear with a very low contribution in the frequency spectrum of the crankshaft's speed (see Fig. 6). On the contrary, if a cylinder generates less work with respect to the others, the first three har-

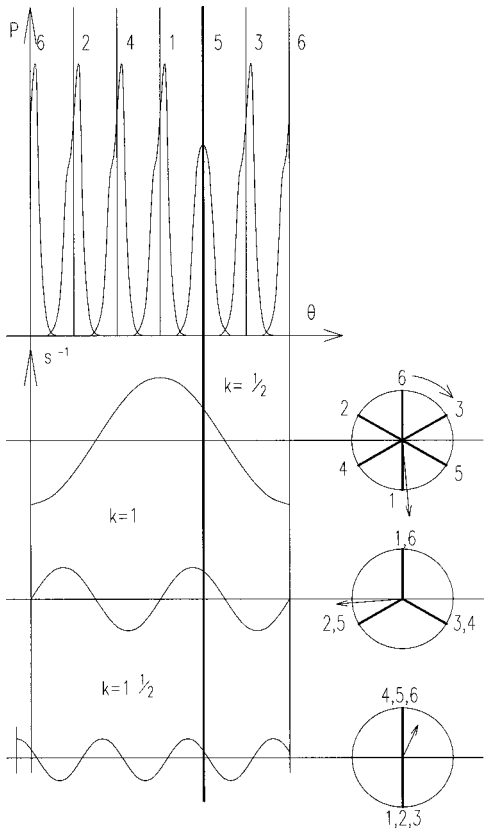


Fig. 9 Identification of a faulty cylinder from the phases of the lowest three harmonic orders of the crankshaft's speed

monic orders have a larger contribution in the frequency spectrum of the gas-pressure torque (see Fig. 5) and determine larger amplitudes of the corresponding components in the frequency spectrum of the crankshaft's speed (see Figs. 6 and 7).

If the frequency spectrum of the crankshaft's speed corresponding to uniform cylinders operation is compared to the spectrum corresponding to a faulty cylinder (see Fig. 6) one may see that the major difference is produced by the amplitudes of the first three harmonic orders. As far as the cylinders operate uniformly, these amplitudes are maintained under a certain limit. Once a cylinder starts to reduce its contribution, the amplitudes of the first three harmonic orders start increasing. These amplitudes may be used to determine the degree by which a cylinder reduces its contribution to the total gas-pressure torque. This situation is illustrated in Fig. 7. If a cylinder does not work, the amplitude of the half harmonic order ($k=1/2$) has almost the same value as the amplitude of the first major order ($k=3$), while if the contribution of a cylinder is reduced by 35 percent with respect to the others, the amplitude of the half order is only about half the amplitude of the first major order, for the same value of the total gas-pressure torque of the engine. A scale could be established in this way to estimate the degree of malfunction.

The identification of the faulty cylinder may be achieved by analyzing the phases of the lowest three harmonic orders.

Figure 9 presents this situation in detail. The upper part of the figure shows the pressure traces of the six cylinders, in a sequence corresponding to the firing order with cylinder #5 disconnected. Underneath, the three harmonic curves, representing the lowest three orders of the measured speed, are drawn respecting the measured amplitudes and phases. It is seen that, only for the expansion stroke of cylinder #5 all three curves have, simultaneously, a negative slope. The phase angle diagrams of these orders are placed on the right-hand side of the figure and, on each one, the vector corresponding to the harmonic component of the measured speed is also represented. One may see that, for each of the three considered orders, the vectors are pointing toward the group of cylinders that produce less work. The cylinder that is identified three times among the less productive cylinders is the faulty one.

Table 1 Identification of the faulty cylinder from the position, in the phase-angle diagrams, of the vectors representing the lowest three harmonic orders of the measured speed

k	CYLINDER					
	#1	#2	#3	#4	#5	#6
1/2	-				-	
1		-			-	
1 1/2				-	-	-

Based on this observation, the following algorithm may be developed.

- The phase-angle diagrams, considering the firing order of the engine (which corresponds to $k=1/2$), are drawn for the lowest three harmonic orders placing in the top dead center (TDC) the cylinder that fires at 0 deg in the considered cycle.
- On these phase angle diagrams, the corresponding vectors of the measured speed are represented in a system of coordinate axes having on the vertical the cosine term and on the horizontal the negative value of the sine term (Fig. 9).
- The cylinders toward which the vectors are pointing are the less contributors and receive a ''-'' mark. If there are cylinders that receive a ''−'' mark for all three harmonic orders they are clearly identified as less contributors to the engine total output. This procedure is presented in Table 1 for the case shown in Fig. 9 (cylinder #5 disconnected).
- This algorithm was able to identify a faulty cylinder as soon as its contribution dropped below 15 percent with respect to the contribution of the other cylinders

Conclusion

1 The DFT of the measured crankshaft's speed, under steady-state operation conditions at constant load shows insignificant variation of the amplitude of the lowest major harmonic order. This is valid both for uniform and nonuniform contributions of the cylinders to the total engine output and the lowest major harmonic order may be used to correlate its amplitude to the IMEP/gas-pressure torque for a given engine speed. This correlation may be stored as a map in the PROM of the engine controls and used to determine, on-line, the IMEP/gas-pressure torque of the engine.

2 The amplitudes of the lowest two harmonic orders ($k=1/2$ and $k=1$) of the measured speed may be used to scale the cylinder power unbalance.

3 An algorithm capable to identify faulty cylinders was developed, based on the phases of the lowest three harmonic orders ($k=1/2$, $k=1$, and $k=1\frac{1}{2}$) of the measured speed.

Acknowledgments

The authors acknowledge the technical support and sponsorship of the Automotive Research Center by the US Army National Automotive Center and TARDEC, Warren, Michigan.

References

[1] Williams, J., ''An Overview of Misfiring Cylinder Engine Diagnostic Techniques Based on Crankshaft Angular Velocity Measurements,'' SAE Paper 960039.
[2] Brown, T. S., and Neil, W. S., ''Determination of Engine Cylinder Pressures from Crankshaft Speed Fluctuations,'' SAE Paper 920463.
[3] Ben-Ari, J., deBotton, G., Itzaki, R., and Sher, E., ''Fault Detection in Internal Combustion Engines by the Vibration Analysis Method,'' SAE Paper 1999-01-1223.
[4] Foerster, J., Lohmann, A., Mezger, M., and Ries-Mueller, K., ''Advanced Engine Misfire Detection for SI-Engines,'' SAE Paper 970855.

[5] Azzoni, P. M., Moro, D., Porceddu-Cilione, C. M., and Rizzoni, G., ''Misfire Detection in a High-Performance Engine by the Principal Component Analysis Approach,'' SAE Paper 960622.
[6] Henein, N. A., Bryzik, W., Taylor, C., and Nichols A., ''Dynamic Parameters for Engine Diagnostics: Effect of Sampling,'' SAE Paper 932411.
[7] Guezenec Y. G., and Gyan Ph., ''A Novel Approach to Real-Time Estimation of the Individual Cylinder Combustion Pressure for S.I. Engine Control,'' SAE Paper 1999-01-0209.
[8] Citron, S. J., O'Higgins, J. E., and Chen, L. Y., ''Cylinder by Cylinder Engine Pressure and Pressure Torque Waveform Determination Utilizing Speed Fluctuation,'' SAE Paper 890486.
[9] Kronberger, M., 1989, ''Einzelzylinder-Regelung beim Fahrzeugdieselmotor—Rechnerunterstutzte Simulation und Versuchergebnisse,'' 10 Internationales Wiener Motorensymposium, Apr. 27–28, VDI Fortschritt-Berichte, Reihe 12 Verkehrstechnik/Fahrzeugtechnik Nr. 122, VDI Verlag.
[10] Iida, K., Akishino, K., and Kido, K., ''IMEP Estimation from Instantaneous Crankshaft Torque Variation,'' SAE Paper 900617.
[11] Chen, K. S., and Chen, S., ''Engine Diagnostics by Dynamic Shaft Measurement: A Progress Report,'' SAE Paper 932412.
[12] Taraza, D., ''Possibilities to Reconstruct Indicator Diagrams by Analysis of the Angular Motion of the Crankshaft,'' SAE Paper 932414.
[13] Champoussin, J. C., and Ginoux, S., ''Engine Torque Determination by Crank Angle Measurements: State of the Art, Future Prospects,'' SAE Paper 970532.
[14] Taraza, D., Henein, A. N., and Bryzik, W., ''Determination of the Gas-Pressure Torque of a Multicylinder Engine from Measurements of the Crankshaft's Speed Variation,'' SAE Paper 980164.
[15] Hafner, K. E., and Maass, H., 1985, *Torsionsschwingungen in der Verbrennungskraftmaschine,* Springer-Verlag, Wien.
[16] Chen, K. S., and Chang, T., ''Crankshaft Torsional and Damping Simulation—An Update and Correlation with Test Results,'' SAE Paper 861226.

Detailed Chemical Kinetic Simulation of Natural Gas HCCI Combustion: Gas Composition Effects and Investigation of Control Strategies

D. Flowers
S. Aceves
C. K. Westbrook
J. R. Smith

Lawrence Livermore National Laboratory,
Livermore, CA 94551

R. Dibble
University of California,
Berkeley, CA 94720

This paper uses the HCT (hydrodynamics, chemistry and transport) chemical kinetics code to analyze natural gas combustion in an HCCI engine. The HCT code has been modified to better represent the conditions existing inside an engine, including a wall heat transfer correlation. Combustion control and low power output per displacement remain as two of the biggest challenges to obtaining satisfactory performance out of an HCCI engine, and these challenges are addressed in this paper. The paper considers the effect of natural gas composition on HCCI combustion, and then explores three control strategies for HCCI engines: DME (dimethyl ether) addition, intake heating and hot EGR addition. The results show that HCCI combustion is sensitive to natural gas composition, and an active control may be required to compensate for possible changes in composition. Each control strategy has been evaluated for its influence on the performance of an HCCI engine. [DOI: 10.1115/1.1364521]

Introduction

Homogeneous charge compression ignition (HCCI) engines are being considered as a future alternative for diesel engines. HCCI engines have the potential for high efficiency (diesel-like [1]), very low nitrogen oxide (NO_x) and particulate emissions, and low cost (because no high-pressure fuel injection system is required). Disadvantages of HCCI engines include: high hydrocarbon (HC) and carbon monoxide (CO) emissions, high peak pressures, high rates of heat release, reduced operating range, low maximum power, difficulty in starting the engine, and difficulty of control.

HCCI was identified as a distinct combustion phenomenon about 20 years ago. Initial papers [2,3] recognized the basic characteristics of HCCI that have been validated many times since then: HCCI ignition occurs at many points simultaneously, with no flame propagation. Combustion was described as very smooth, with very low cyclic variations. Noguchi et al. [3] also conducted a spectroscopic study of HCCI combustion. Many radicals were observed, and they were shown to appear in a specific temporal sequence. In contrast, with spark-ignited (SI) combustion all radicals appear at the same time (probably distributed in the same spatial sequence through the flame front). These initial experiments were done in 2-stroke engines, with very high EGR.

Since then, HCCI two-stroke engines have been developed to the point of commercialization for motorcycles [4]. HCCI motorcycle engines have higher fuel economy, lower emissions and smoother combustion than 2-stroke spark-ignited engines. However, HC and CO emissions out of the HCCI engine are still very high compared with the current automotive emissions standards. An improved version of the engine has been recently evaluated [5], which shows improvements in fuel economy and emissions.

Najt and Foster [6] did the first HCCI experiment with a four-stroke engine. They also analyzed the process, considering that HCCI is controlled by chemical kinetics, with negligible influence from physical effects (turbulence, mixing). Najt and Foster used a simplified chemical kinetics model to predict heat release as a function of pressure, temperature, and species concentration in the cylinder.

Since then, a description of the HCCI process has gained acceptance. HCCI has been described as dominated by chemical kinetics, with little effect of turbulence. Crevices and boundary layers are too cold to react, and result in hydrocarbon and CO emissions. Combustion at homogeneous, low equivalence ratio conditions results in modest temperature combustion products, which do not generate NO_x or particulate matter.

Physical understanding has not yet resulted in a solution to the problems of operating a four-stroke engine in HCCI mode. The control issue appears to be most important. Some alternatives have been described [7,8], but further work is required to identify a general control strategy.

Analysis of HCCI engines is not well developed, even though the process may be reasonably well understood. Most publications on HCCI present only experimental results. Of those that present analysis, some have used a fluid mechanics code [9] with a very simplified chemical kinetics model. The use of fluid mechanics codes is appropriate in operating conditions such as PREDIC (early direct injection [10]), where the charge is not homogeneous and fuel mixing and evaporation may have a significant effect in the combustion process. For homogeneous charge engines, the process is mainly dominated by chemistry, and it is more important to have a detailed chemical kinetics model than a fluid mechanics model. Therefore, valuable predictions and results can be obtained from single-zone chemical kinetics analyses that assume that the combustion chamber is a well-stirred reactor with uniform temperature and pressure [11,12]. This is the tool used in this paper. The ideal tool for HCCI analysis is a combination of a fluid mechanics code with a detailed chemical kinetics code. This is, however, beyond our current computational capabilities.

The concept considered here is a high compression ratio, lean burn, natural gas engine with homogeneous charge compression ignition (HCCI). This engine has the potential to achieve high efficiency and very low NO_x emission. The key to achieving this potential is developing control methods that allow for consistent

Contributed by the Internal Combustion Engine Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the ICE Division Apr. 25, 2000; final revision received by the ASME Headquarters Dec. 6, 2000. Associate Editor: D. Assanis.

operation over the range of operating conditions. Because HCCI combustion is an autoignition process, heat release occurs very rapidly. The major factors that contribute to the initiation of this autoignition process in an engine cycle are pressure, temperature, fuel composition, equivalence ratio, and engine speed. These factors are influenced by many operating parameters such as inlet heating, supercharging, and residual gas trapping. The control methods are designed to manage the heat release process to occur at the appropriate time in the engine cycle.

Modeling of the HCCI Combustion Process

A zero-dimensional detailed kinetic model is used here to model HCCI combustion. A premixed charge at uniform temperature and pressure is compressed and expanded at a rate that depends on the engine speed and geometry. This kind of model cannot capture the multi-dimensional processes that occur in a real engine cylinder, but, since the heat release is a global nonpropagating autoignition process, a zero-dimensional model can reasonably capture the start of combustion and heat release of the core mixture. Since the start of combustion of the central core dictates the overall process, control of this combustion timing will control performance.

All of the modeling computations in this study were carried out using the HCT code (hydrodynamics, chemistry and transport [13]). This code has been used in a large number of investigations over the years, and in particular was used in past studies of engine knock and autoignition [14–16]. HCT permits the use of a variety of boundary and initial conditions for reactive systems, depending on the needs of the particular system being examined. In the present case, the relevant conditions are those which describe the bulk gases in the combustion chamber.

Autoignition of a homogeneous charge is virtually identical to the knock process that can occur in the Otto cycle engine. From detailed kinetic modeling of motored engines using a homogeneous charge of a variety of fuels, it is known that the controlling parameters in the initiation of this process are the fuel components (mixtures behave differently than neat fuels), and the temperature and density history that the fuel-air mixture experiences. Motored engine experiments at General Motors Research Laboratories [15,17] agree well with HCT simulation of this process. Thus if the precise conditions at the start of compression are known (species, temperature and pressure), the beginning of combustion can be accurately predicted.

During an engine cycle, a number of processes occur which influence the time variation of the temperature and pressure of the bulk reactive gases in the combustion chamber. Piston motion first compresses and thus heats the bulk gases and then expands and cools them. During this time chemical reactions release heat and change the overall composition of the gases. Fresh unreacted fuel and air are added to hot residual gases left over from the previous cycle. Residual gases from previous engine cycles, which consist largely of water vapor, CO_2, and molecular nitrogen and oxygen are assumed to become fully mixed with the fresh charge. In addition, heat losses to the engine chamber walls, blowby, fuel trapping in crevice volumes, and other processes occur. In the current simulations only heat transfer losses are taken into account.

The computational model treats the combustion chamber as a homogeneous reactor with a variable volume. The mixed temperature of the residual gases and the fresh charge is estimated by a published procedure [18]. The volume is changed with time using a slider-crank formula. The heat transfer submodel employed in the HCT code simulations uses Woschni's correlation [19].

The present analysis considers a single zone, lumped model that ignores spatial variations in the combustion chamber, treating heat loss as a distributed heat transfer rate, proportional to the temperature difference between the average gas temperature and a time-averaged wall temperature. The authors recognize that this is a great oversimplification of the actual condition within the combustion chamber. In particular, the boundary layer, which contains significant mass, must be at a lower temperature than the bulk gas near TDC. Due to the assumed temperature uniformity our estimates of burn duration and the heat release processes will be shorter than in experiments. That is, the boundary layer and crevices will always react last and extend the heat release rate compared to this simulation. Peak cylinder pressure and rate of pressure rise are thus overestimated with the current single-zone model, and the model cannot accurately predict CO and hydrocarbon emissions, which primarily depend on poor combustion in the boundary layer and crevices. Predictions of start of combustion and NO_x, which depend on the peak temperature of the core gases inside the cylinder, have been shown to be determined with reasonable accuracy [12].

Two reaction mechanisms are used in this study. The first mechanism includes species through C_4 [17], and models natural gas autoignition chemistry. The second mechanism is a reduced set with species through C_2 and additionally including dimethyl ether (DME) oxidation chemistry [20]. Both mechanisms include NO_x kinetics from the Gas Research Institute mechanism version 1.2 [21]. The chemical kinetic reaction mechanisms used by the model for natural gas ignition and NO_x production have been extremely well established and are widely used.

The simulation is started at BDC where fresh charge at specified pressure, temperature and composition is inducted into the cylinder. If EGR is used, the proper fraction of major residual gas components is added. The cylinder wall, piston, and head are all assumed to be at a uniform 430 K. Chemical reactions are computed explicitly in the kinetics model. Thermodynamic table values of c_p and c_v are explicitly calculated to account for enthalpy and pressure changes. The simulation is stopped at BDC at the end of the expansion stroke and the indicated efficiency is computed. The NO_x values reported are taken at BDC.

Results

Simulations have been performed to look at the effect of several different control parameters on HCCI combustion. One study uses the detailed natural gas chemistry model to look at the effect of natural gas composition on the peak of heat release in an HCCI engine. Because of the rapid heat release process, peak of heat release typically occurs within 1–2 CAD of start of combustion. The effect of DME addition, inlet heating, and EGR has also been evaluated with respect to the control of combustion timing.

In these simulations the ''best'' operating point for an HCCI engine is taken to be peak heat release rate at TDC. This choice is debatable. Figure 1 shows experimental results of gross indicated efficiency versus combustion timing for HCCI engines, adapted from [11]. The combustion timing is specified by the peak rate of heat release. The data represents a wide variety of fuels under different operating conditions. According to these data, maximum efficiency occurs at slightly advanced combustion timing. A tradeoff exists between NO_x and HC emissions relative to combustion timing. As combustion timing advances, NO_x emissions rapidly increase, and delaying combustion results in rapid increase in HC emissions [11]. The best combustion timing depends on a variety of factors and likely varies depending on the operating conditions. The use of TDC combustion timing in this study seeks to address the trade-offs in efficiency, NO_x, and HC emissions at some level. The selection of TDC combustion as ''best'' timing provides a baseline for comparison of different control strategies and timing effects due to fuel composition

Effect of Fuel Composition on Combustion Timing. The HCT engine code has been used to assess the effect of variations in natural gas fuel composition on the start of combustion in an HCCI operated engine. The engine selected for these simulations is a 1.9 L engine based on the Volkswagen TDI. The TDI has a bore of 79.5 mm and a stroke of 95.5 mm. The model engine has a compression ratio of 18, and the absolute intake pressure is

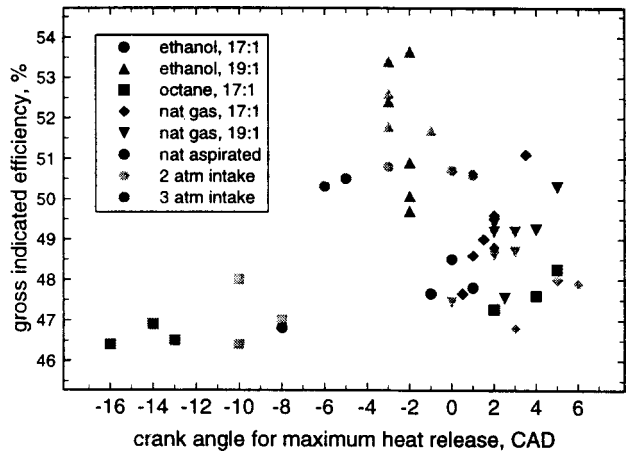


Fig. 1 Gross indicated thermal efficiency versus timing for peak heat release adapted from Lund Institute data [11]

assumed equal to 2.0 bar. This engine has been selected because it is an appropriate geometry for a small vehicular engine, and will be studied in future experimental research.

These fuel sensitivity simulations compare crank angle for maximum heat release of a natural gas blend relative to a baseline blend operated in the same conditions. Because methane is the major component of natural gas, pure methane is the baseline fuel. The BDC temperature is adjusted so that the peak heat release of pure methane, for a given equivalence ratio and engine speed, occurs at TDC. The engine is run with no EGR and the BDC temperature at the beginning of the compression stroke is specified.

Figure 2 shows the results of methane combustion with maximum heat release at TDC for three equivalence ratios with varying engine speed. The figure shows that the BDC temperature for start of combustion must increase with increasing engine speed. The shorter reaction time available to achieve combustion at higher speeds makes it necessary to start the process at a higher temperature for obtaining peak heat release at TDC.

Natural gas is typically a mixture of methane, ethane, propane, and butane, as well as small amounts of other higher hydrocarbons and inert species (usually less than 1 percent–2 percent by volume). Regional and seasonal factors play a significant role in the specific composition that a producer delivers. The Gas Research Institute studied the natural gas composition in the United States [22]. Higher hydrocarbons in the fuel cause autoignition to start earlier than for pure methane. Low-temperature reactions occur due to the higher hydrocarbons that do not occur in pure

Table 1 Fuel blends used for sensitivity analysis study

mol% CH_4	mol% C_2H_6	mol% C_3H_8	mol% C_4H_{10}	Notes
100.00	0.00	0.00	0.00	
96.50	3.50	0.00	0.00	Methane/Ethane Blend
93.00	7.00	0.00	0.00	Methane/Ethane Blend
86.00	14.00	0.00	0.00	Methane/Ethane Blend
98.30	0.00	1.70	0.00	Methane/Propane Blend
96.60	0.00	3.40	0.00	Methane/Propane Blend
93.20	0.00	6.80	0.00	Methane/Propane Blend
99.30	0.00	0.00	0.70	Methane/Butane Blend
98.60	0.00	0.00	1.40	Methane/Butane Blend
97.20	0.00	0.00	2.80	Methane/Butane Blend
94.10	3.50	1.70	0.70	
88.20	7.00	3.40	1.40	
76.40	14.00	6.80	2.80	
87.15	7.63	3.70	1.53	
84.18	14.59	0.90	0.34	
60.43	2.01	37.30	0.26	Extreme Peakshaving Blend
93.14	4.61	2.05	0.20	Massachusetts City 2
83.43	1.94	14.40	0.23	Maryland Propane-Air Peakshaving
95.85	3.44	0.51	0.20	Virginia Station A
93.53	4.69	1.36	0.42	Ohio
92.92	4.72	1.64	0.72	Texas City #1
91.06	7.62	1.10	0.22	Colorado Station A
94.21	4.24	1.14	0.41	California City #1 Gate Station A

methane, causing additional charge heating and introduction of radicals well before TDC. In order to understand the effect of composition on the ignition process, several compositions of natural gas are tried and the advance of maximum heat release from TDC relative to pure methane combustion is determined using HCT. Using the Gas Research Institute data, several fuel blends have been selected that define a large range in composition. The natural gas is specified to consist only of the four primary components: methane, ethane, propane, and butane. Inert species are neglected as part of the composition because they add unnecessary complexity with little effect on the results. Also neglected are hydrocarbons higher than butane, which typically are less than 0.1 mole percent of the gas composition. Table 1 presents the gas blends used for these studies.

Figure 3 shows the advance in crank angle for maximum heat release for the different natural gas blends. For a fixed equivalence ratio and engine speed the BDC temperature that results in maximum heat release at TDC for methane is determined. Using the methane BDC temperature the crank angle at which maximum heat release occurs for the different natural gas compositions is determined. The advance is plotted as a function of a figure of merit (FOM), which was found to collapse the data into a single curve. The figure of merit is calculated by scaling the mole fraction of ethane by a factor of 1/6 and adding that to the mole fraction of propane and butane. This 1/6 factor for the ethane mole fraction was determined heuristically by assessing what scaling

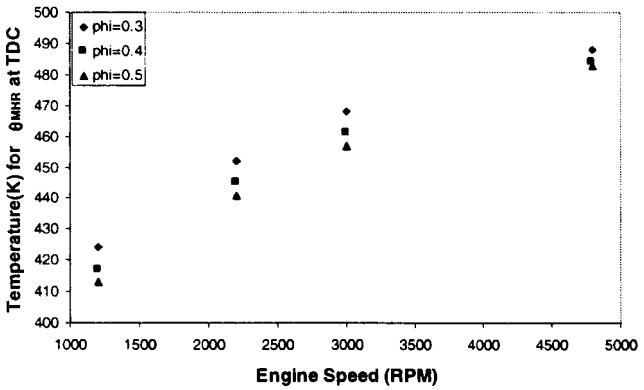


Fig. 2 Temperature for TDC peak heat release of pure methane versus engine speed

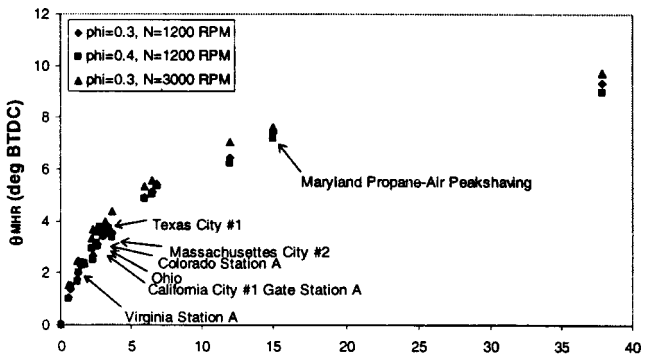


Fig. 3 Crank angle advance of peak heat release versus fuel composition figure of merit for fuel blends

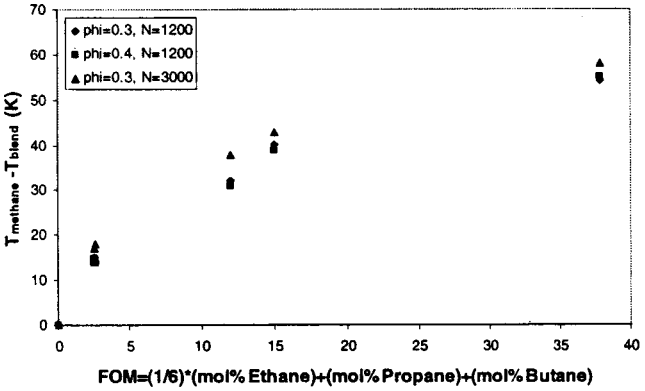


Fig. 4 Temperature for TDC peak heat release versus figure of merit for various fuel blends

resulted in the data falling on a single curve with the least scatter. This scaling implies that the dominant components resulting in early combustion are propane and butane. The 23 natural gas blends of Table 1 have been evaluated at each of the three equivalence ratio/engine speed combinations. The important trend that these simulations show is the high sensitivity of heat release timing that can be seen for figures of merit up to 5, a range that represents very typical blends delivered by gas suppliers.

A second study has been performed to determine what BDC temperature is required for different gas mixtures to achieve maximum heat release at TDC. Six blends from the previous 23 have been selected as extremes in composition of each of the components as well as the overall blends. The reduction in BDC temperature with respect to methane required to achieve maximum heat release at TDC for each of the blends is shown in Fig. 4, as a function of the same figure of merit used in Fig. 3. The required temperature reduction is fairly consistent for the three cases.

Figures 3 and 4 indicate that compensation will be required to account for normal variations in fuel composition across the U.S., and a closed loop control system with an indicator of the timing of heat release may be necessary to account for these variations. It should be noted that the blends creating the most advance are not typical and would likely occur only during ''peak shaving'' when a significant amount of propane is blended into the fuel during heavy natural gas use periods [22].

Control Methods for HCCI Engines. Three different methods have been studied for their ability to control combustion timing in HCCI engines: DME addition, intake heating, and EGR. The engine modeled in these simulations is a Cooperative Fuels Research engine with a bore of 82.5 mm, a stroke of 114.3 mm, and a connecting rod length of 254 mm. Two compression ratios are considered: 12:1 and 18:1, with natural aspiration and 1 bar supercharge (2 bar absolute intake manifold pressure). Engine speed is fixed at 1800 rpm. The methane/DME reaction mechanism is used for these calculations. The control variable is adjusted so that the peak heat release always occurs at TDC. This engine and fuel have been selected to complement experimental research currently underway at UC Berkeley. Operating conditions have been selected based on experimental parameters. These results should always be taken as relative indicators of the performance of the engine with respect to the control parameters.

It is considered that operation of the engine is constrained by peak cylinder pressure and NO_x emissions. The maximum peak cylinder pressure is set at 250 bars. This value is higher than the existing limit in most current engines. However, previous validation of the model has shown that single-zone HCT analysis overpredicts peak cylinder pressure, and it is therefore reasonable to use a high value as a limit [12]. The NO_x emissions are restricted

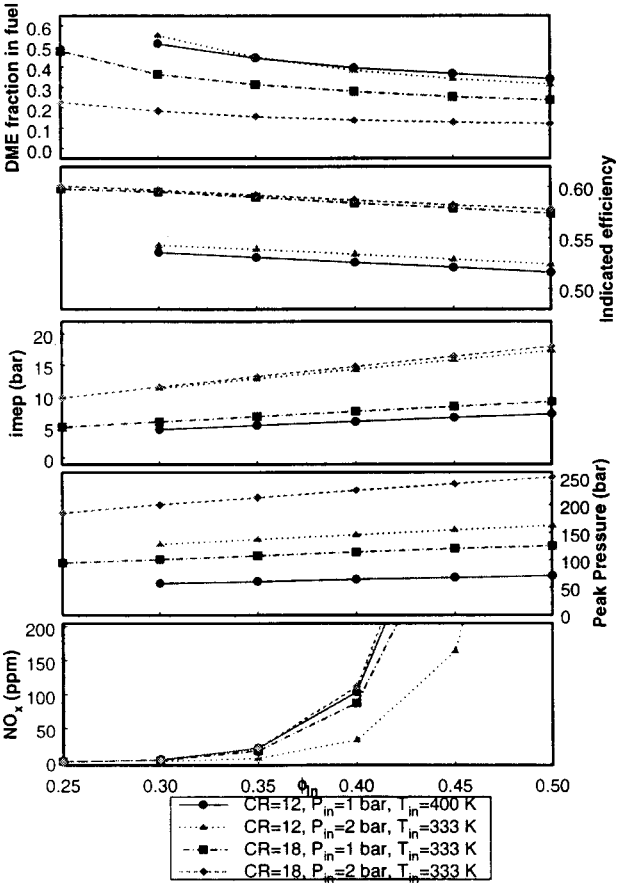


Fig. 5 Simulation results for HCCI control by DME addition

to 100 ppm. This limit would result in very low driving cycle emissions if the engine were used in an automobile.

Dimethyl Ether (DME) Additive Control. Autoignition timing may be controllable by blending low cetane number fuel (natural gas) with high cetane number fuel. Dimethyl ether is an ideal fuel additive for natural gas HCCI because it has a short ignition delay and has well characterized reaction chemistry [20].

Figure 5 shows the results for the addition of DME to the engine. DME fraction in the fuel, indicated efficiency, indicated mean effective pressure, peak cylinder pressure, and NO_x are plotted versus intake equivalence ratio for the four different conditions. DME concentration is always adjusted to achieve peak heat release at TDC. Residual gases are not considered in this analysis so the charge is pure fuel and air and therefore the intake equivalence ratio is equal to the in-cylinder equivalence ratio. The inlet temperature is fixed at 333 K for each case except for the naturally aspirated 12:1 compression ratio. In that case the inlet temperature is increased to 400 K because satisfactory combustion could not be achieved for the range of equivalence ratio without some inlet heating. Satisfactory combustion refers to having a significant fraction (say 80 percent) of the fuel burned. In practice it is very easy to determine complete combustion in the single zone model because either very little fuel reacts or nearly all reacts. The equivalence ratio reported includes both DME and methane. Complete combustion could not be achieved for the 12:1 compression ratio cases with equivalence ratio below 0.3, and could not be achieved for the 18:1 compression ratio cases below equivalence ratio of 0.25.

The maximum values of the performance variables achievable during operation may be constrained by peak cylinder pressure and a specified limit on NO_x. As Fig. 5 shows, only the super-

charged 18:1 compression ratio case approaches the 250 bar peak pressure limit. In all the cases analyzed, the NO_x threshold is the governing constraint. If greater intake boost is used the pressure limit will likely play a stronger role. The NO_x constraint limits both naturally aspirated cases and the supercharged 18:1 compression ratio case to an equivalence ratio of 0.4, while the maximum equivalence ratio of the supercharged 12:1 compression ratio case is roughly 0.43. The DME fraction required for the 12:1 cases ranges from slightly over 50 percent at low load to 30 percent–35 percent at high load. For 18:1 compression ratio, the naturally aspirated engine requires 50 percent DME at low load to 30 percent at high load, and the supercharged engine requires 25 percent at low load to 15 percent at high load.

The gross indicated efficiencies range from 58 percent to 60 percent for 18:1 compression ratio, and from 52 percent to 54 percent for the 12:1 compression ratio cases. The indicated efficiency decreases with increasing equivalence ratio due to decreasing specific heat ratio and increasing heat transfer as the equivalence ratio increases. These efficiency results indicate the potential for high efficiency operation of HCCI engines.

The imep, in the allowable range of operation, ranges from 10 to 15 bars for the 18:1 compression ratio supercharged case, and from 11 to 15 bars for the 12:1 compression ratio supercharged case. For the 18:1 compression ratio naturally aspirated case the imep ranges from 5 to 7.5 bars, and ranges from 5 to 6 bars for the 12:1 compression ratio naturally aspirated case. The naturally aspirated maximum imep results are fairly low, and indicate that supercharging is necessary to increase the power density. Because the lowest calculated imep achievable in this control method is 5 bars, throttling may be required for idle operation, to balance a typical 1–2 bar frictional loss at these conditions [18].

Inlet Heating Control. HCCI control by inlet heating involves adjusting the mixture temperature so that conditions are appropriate for autoignition of the charge to occur at the desired crank angle. A procedure for heating the intake gases, such as an electric heater, an air-to-air heat exchanger with hot exhaust gases or a turbocharger would be required in an operating engine. Figure 6 shows inlet temperature, indicated efficiency, imep, peak pressure, and NO_x plotted versus equivalence ratio for the CFR engine operated with inlet heating to obtain maximum heat release at TDC. The mixture at the start of the cycle is fuel and air only with no residual gases considered. The inlet temperature is the temperature in the cylinder at beginning of the compression stroke.

The 250 bar pressure constraint is not approached in any of the simulations. The 100 ppm NO_x constraint is reached at equivalence ratios of 0.31 to 0.32 for all cases. This NO_x limit could probably be extended with some EGR, at the expense of efficiency. Supercharging and increasing the compression ratio decrease the required inlet temperature, and this appears to also reduce the peak temperature at a fixed equivalence ratio, leading to lower NO_x for these cases. For the 12:1 compression ratio naturally aspirated engine the simulation results indicate temperatures of about 550 K to obtain combustion at TDC. Supercharging a 12:1 compression ratio engine to 2 bar intake pressure reduces the required temperature down to the range of 525–540 K. For the 18:1 compression ratio naturally aspirated case the inlet temperature required for TDC combustion is in the range of 465–470 K, and for the 18:1 supercharged case the inlet temperature range is 435–450 K. With inlet heating, auto-ignition can be achieved even at very low equivalence ratio. The compression done by supercharging heats the intake air, so that little additional heating may be required under some supercharged conditions. It should also be noted that, in some cases as equivalence ratio is increased, higher temperatures are required to achieve TDC peak heat release, due to decreasing specific heat ratio and chemistry effects.

The indicated efficiency ranges from 55 percent to 60 percent for the 18:1 compression ratio cases and from 49 percent to 54 percent for the 12:1 compression ratio cases, with the two supercharged cases having slightly higher efficiency over most of the range than the respective naturally aspirated cases. Again, these efficiency results demonstrate the potential for very high efficiency in an HCCI engine.

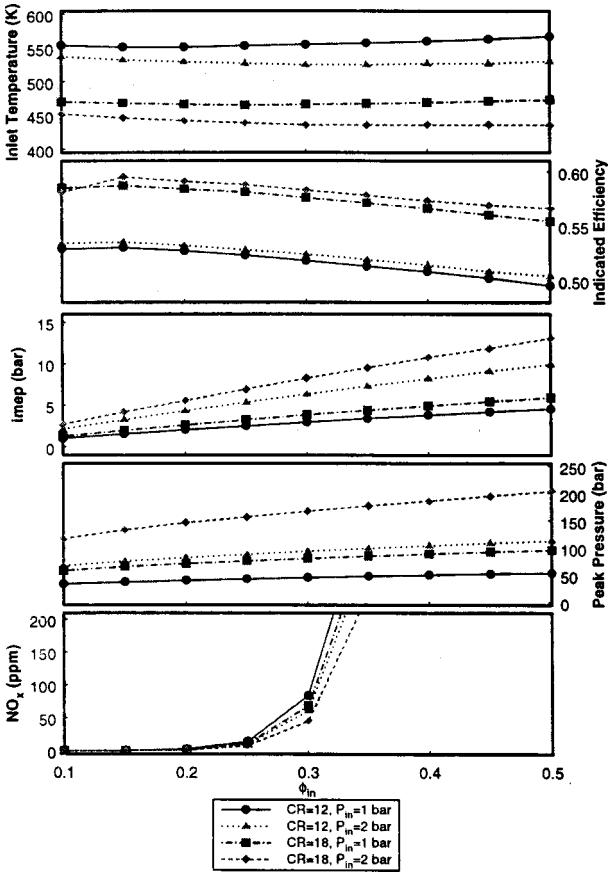


Fig. 6 Simulation results for HCCI control by intake heating

For the constrained range of operation, the imep ranges from 3 to 8 bars for the supercharged 18:1 compression ratio, 1.3 to 4 bars for the naturally aspirated 18:1 compression ratio, 2 to 6 bars for the supercharged 12:1 compression ratio, and 1 to 3 bars for the naturally aspirated 12:1 compression ratio. The peak numbers suggest that even more supercharging may be required to achieve reasonable power density out of an HCCI engine. This could lead to the peak cylinder pressure being the dominant constraint. The low-load imep results show that frictional work can be balanced by the combustion work without throttling.

EGR Control. Blending the intake fuel-air mixture with hot exhaust gases has several effects. First, the exhaust gases will raise the charge temperature. Second, the specific heat ratio of the mixture will decrease resulting in lower TDC temperature and pressure relative to a pure fuel/air mixture. In a lean burn engine, the intake fuel-air ratio, based on the mixture of fuel and air before addition of exhaust gases, will be different than the in-cylinder fuel-air ratio when exhaust gas has been blended with the intake charge. The relation below accounts for the additional air present in the cylinder due to addition of exhaust gas in a lean combustion system:

$$\phi_{\text{act}}=\frac{(1-\text{EGR})\phi_{\text{in}}}{1-\text{EGR}\phi_{\text{in}}} \tag{1}$$

In Eq. (1), EGR is the fraction of exhaust gas in the cylinder relative to the inducted fresh fuel and air, ϕ_{in} is the intake equivalence ratio, and ϕ_{act} is the in-cylinder equivalence ratio. In an operating engine hot exhaust gases can be introduced into the cylinder in several ways. For example, an insulated line could be

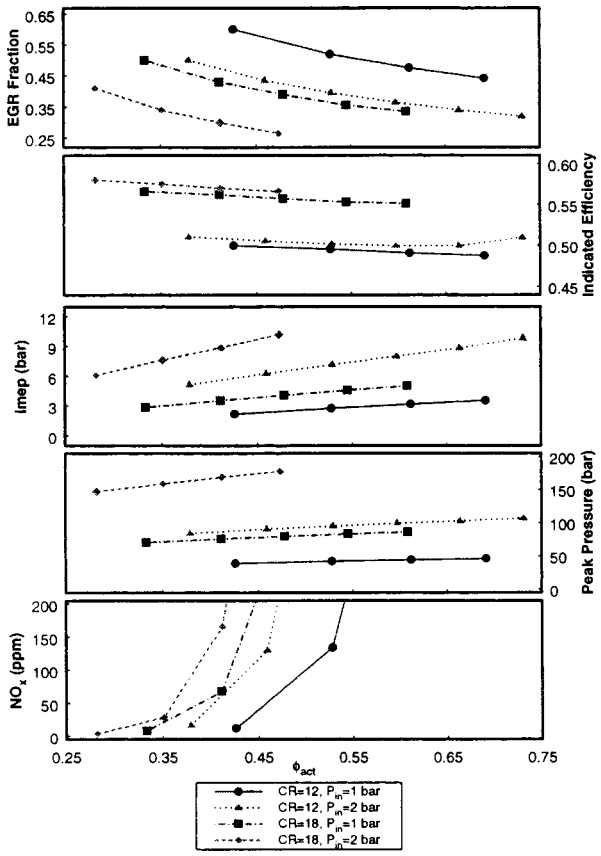


Fig. 7 Simulation results for HCCI control by EGR

run from the exhaust and blended with the intake air or the engine exhaust value could be closed early with variable valve timing.

Figure 7 shows the EGR fraction, indicated efficiency, imep, peak cylinder pressure, and NO_x plotted versus the in-cylinder equivalence ratio. The maximum cylinder pressure constraint is not met in any of these cases. The NO_x limit governs the operation range. The maximum equivalence ratio allowable based on the NO_x constraint ranges between 0.5 for the naturally aspirated 12:1 compression ratio and 0.39 for the supercharged 18:1 compression ratio. The lower NO_x emissions for a given equivalence ratio compared with the results for preheated intake (Fig. 6) are due to the lower specific heat of the residual gases, which reduces the peak cylinder temperature in the EGR case. The EGR fraction required for TDC peak heat release decreases with supercharging and increased compression ratio. The EGR fraction ranges from 60 percent for low load in the naturally aspirated 12:1 case to 26 percent in the 18:1 case at high load. The left-hand end of each curve on the EGR fraction versus equivalence ratio plot is the lowest equivalence ratio at which complete combustion at TDC could be achieved. Supercharging and increased compression ratio both have the affect to extend the operating limit to leaner mixtures.

Gross indicated efficiency is higher than 49 percent in all cases, and reaches 58 percent for the supercharged 18:1 case. The imep ranges from 2 to 2.5 bars for the naturally aspirated 12:1 compression ratio, from 5 to 6 bars for the supercharged 12:1 compression ratio, from 3 to 3.7 bars for the naturally aspirated 18:1 compression ratio, and from 6 to 8 bars for the supercharged 18:1 compression ratio. Further supercharging or relaxation of the NO_x limit at peak load may be necessary to achieve reasonable power density. The low-load imep results indicate that some throttling may be necessary for idle operation.

Discussion of Control Methods. Each of the control methods analyzed in this paper has the potential to control combustion timing in an HCCI engine. Some general observations can be made about how these methods can be applied to HCCI engine control. Supercharging will be required to achieve reasonable power density. The intake air compression method, either with a mechanical supercharger or turbocharger, will need to be able to accommodate a wide range of pressure ratios, possibly up to 2 bars of pressure boost. The mechanical pressure limits of the engine will be a more important constraint with higher levels of supercharging. The constraint on peak NO_x also affects power density. Boosting the intake pressure and increasing the compression ratio decreases the quantity of a particular control variable necessary to achieve TDC combustion, i.e. less DME additive, less inlet heating, or less EGR is required. Higher compression ratio engines appear to be the best choice for an HCCI engine, particularly due to the inherent higher efficiency, although the lower compression ratio engines may better handle very low loads. Lower compression ratio engines would probably yield higher imep, considering peak pressure and NO_x limits. Previous experimental work has shown that late combustion timing allows for reducing NO_x emissions at higher equivalence ratio at the expense of lower combustion efficiency and greater HC and CO emissions [23]. The peak cylinder pressure may also be reduced for a given operating condition by going to later combustion timing. The tradeoff between lower NO_x and increased HC and CO emissions as timing occurs later may be worthwhile to extend the operating range. Increasing heat transfer is a very important factor for determining the upper limit of compression ratio as the crevice height becomes shorter and more of the mixture is in the boundary layer. A variable compression ratio engine could allow for a wide range of loads without throttling. A fixed compression ratio machine with inlet valve timing control could be used to effectively vary the compression ratio, and with inlet and exhaust valve timing control the intake and exhaust could be effectively throttled without pumping losses.

Inlet heating allows for the greatest control of the operating range particularly at low load. Extremely lean mixtures can be ignited and very low power output can be produced. However, NO_x rises very rapidly at relatively low equivalence ratio, limiting the low emission operating range. Implementation of inlet heating could require additional energy cost to power an external heater, particularly for startup. In an automotive application, requiring the operator to wait for the intake air to heat up could be an unacceptable inconvenience. A DME additive scheme has potential for rapid startup because the fuel composition could be adjusted rapidly to create conditions necessary for startup. Using DME additive control may not be practical over the range of operation because it would require large quantities of both natural gas and DME to be stored. Also the DME results showed a significant amount of throttling required for idle. If used only for startup, only a small storage volume would be required and it could be sized for refill at a maintenance interval. An EGR-controlled engine could probably achieve reasonable peak load power with the advantage that the exhaust gas is readily available in a warmed up engine, but, as with inlet heating, a startup scheme is necessary.

Conclusions

This paper presents an analysis of natural gas HCCI combustion using a detailed chemical kinetics code adapted to simulate engine conditions. Two main issues are analyzed: the effect of natural gas composition on HCCI combustion; and the use of three control methodologies for obtaining satisfactory and reliable HCCI combustion. The three control methodologies used are: addition of DME, intake gas preheat, and mixing with hot EGR. The conclusions from the analysis are:

- HCCI combustion is sensitive to natural gas composition, and an active control will be required to compensate for possible changes in composition typical in the U.S. Changes in natural gas composition may shift the peak heat release timing by as much as

10 crank angle degrees. This change would have a significant effect on emissions, peak cylinder pressure, and efficiency.

• A figure of merit has been obtained that represents to a good approximation the effect of higher hydrocarbons on HCCI combustion. The figure of merit is dominated by the mole fraction of propane and butane.

• The three control strategies being considered have a great effect in changing the combustion parameters for the engine, and should be able to control HCCI combustion. Optimum HCCI combustion (defined as combustion that results in maximum indicated efficiency) can be maintained over a wide range of operating conditions with these control strategies.

• The results indicate a low power output. Supercharging beyond current values (2 bars absolute) is necessary to achieve values comparable to those obtained with naturally aspirated spark-ignited engines.

Acknowledgments

This project is funded by DOE Office of Advanced Automotive Technologies, Steve Chalk and Lucito Cataquiz, Program Managers, and the DOE Office of Heavy Vehicle Technologies, Steve Gougen and Gurpreet Singh, Program Managers. This work was performed under the auspices of the U.S. Department of Energy by the University of California Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48.

References

[1] Suzuki, H., Koike, N., Ishii, H., and Odaka, M., 1997, ''Exhaust Purification of Diesel Engines by Homogeneous Charge with Compression Ignition Part 1: Experimental Investigation of Combustion and Exhaust Emission Behavior Under Pre-Mixed Homogeneous Charge Compression Ignition Method,'' SAE paper 970313.

[2] Onishi, S., Jo, S. H., Shoda, K., Jo, P. D., and Kato, S., 1979, ''Active Thermo-Atmosphere Combustion (ATAC)—A New Combustion Process for Internal Combustion Engines,'' SAE paper 790501.

[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 790840.

[4] Ishibashi, Y., and Asai, M., 1996, ''Improving the Exhaust Emissions of Two-Stroke Engines by Applying the Activated Radical Concept,'' SAE paper 960742.

[5] Ishibashi, Y., and Asai, M., 1998, ''A Low Pressure Pneumatic Direct Injection Two-Stroke Engine by Activated Radical Combustion Concept,'' SAE paper 980757.

[6] Najt, P. M., and Foster, D. E., 1983, ''Compression-Ignited Homogeneous Charge Combustion,'' SAE paper 830264.

[7] Smith, J. R., Aceves, S. M., Westbrook, C. and Pitz, W., 1997, ''Modeling of Homogeneous Charge Compression Ignition (HCCI) of Methane,'' Proceedings of the 1997 ASME Internal Combustion Engine Fall Technical Conference, paper No. 97-ICE-68, ICE-Vol. 29-3, pp. 85–90.

[8] Willand, J., Nieberding, R.-G., Vent, G., and Enderle, C., 1998, ''The Knocking Syndrome—Its Cure and Potential,'' SAE paper 982483.

[9] Amsden, A. A., 1993, ''KIVA-3: A KIVA Program with Block-Structured Mesh for Complex Geometries,'' Los Alamos National Laboratory Report LA-12503-MS.

[10] Miyamoto, T., Hayashi, A. K., Harada, A., Sasaki, S., Hisashi, A., and Tujimura, K., 1999, ''A Computational Investigation of Premixed Lean Diesel Combustion,'' SAE paper 1999-01-0229.

[11] Christensen, M., Johansson, B., Amneus, P., and Mauss, F., 1998, ''Supercharged Homogeneous Charge Compression Ignition,'' SAE paper 980787.

[12] Aceves, S., Smith, J. R., Westbrook, C, and Pitz, W., 1999, ''Compression Ratio Effect on Methane HCCI Combustion,'' ASME Journal of Gas Turbines and Power, Vol. 121, No. 3, July 1999.

[13] Lund, C. M., 1978 ''HCT—A General Computer Program for Calculating Time-Dependent Phenomena Involving One-Dimensional Hydrodynamics, Transport, and Detailed Chemical Kinetics,'' Lawrence Livermore National Laboratory Report UCRL-52504.

[14] Westbrook, C. K., Pitz, W. J., and Leppard, W. R., 1991, ''The Autoignition Chemistry of Paraffinic Fuels and Pro-Knock and Anti-Knock Additives: A Detailed Chemical Kinetic Study,'' SAE paper 912314.

[15] Pitz, W. J., Westbrook, C. K., and Leppard, W. R., 1991, ''Autoignition Chemistry of C4 Olefins Under Motored Engine Conditions: A Comparison of Experimental and Modeling Results,'' SAE paper 912315.

[16] Westbrook, C. K., Warnatz, J., and Pitz, W. J., 1988, ''A Detailed Chemical Kinetic Reaction Mechanism for the Oxidation of iso-Octane and n-Heptane over an Extended Temperature Range and its Application to Analysis of Engine Knock,'' Twenty-Second Symposium (International) on Combustion, p. 893, The Combustion Institute, Pittsburgh.

[17] Curran, H. J., Gaffuri, P., Pitz, W. J., Westbrook, C. K., and Leppard, W. R., 1995, ''Autoignition Chemistry of the Hexane Isomers: An Experimental and Kinetic Modeling Study,'' SAE paper 952406.

[18] Heywood, J. B., 1988, *Internal Combustion Engine Fundamentals*, McGraw-Hill, New York.

[19] Woschni, G., 1967, ''Universally Applicable Equation for the Instantaneous Heat Transfer Coefficient in the Internal Combustion Engine,'' SAE paper 670931.

[20] Curran, H. J., Pitz, W. J., Marinov, N. M., and Westbrook, C. K., 1997, ''A Wide Range Modeling Study of Dimethyl Ether Oxidation,'' International Journal of Chemical Kinetics, Vol. 30, No. 3, pp. 229–241, March 1998.

[21] Frenklach, M., Wang, H., Goldenberg, M., Smith, G. P., Golden, D. M., Bowman, C. T., Hanson, R. K., Gardiner, W. C., and Lissianski, V., 1995, ''GRI-Mech—An Optimized Detailed Chemical Reaction Mechanism for Methane Combustion,'' GRI Topical Report No. GRI-95/0058.

[22] Liss, W. E., Thrasher, W. H., Steinmetz, G. F., Chowdiah, P., and Attari, A., 1992, ''Variability of Natural Gas Composition in Select Major Metropolitan Areas of the United States,'' GRI Report No. GRI-92/0123.

[23] Christensen, M., and Johansson, B, 1998, ''Influence of Mixture Quality on Homogeneous Charge Compression Ignition,'' SAE paper 982454.

Biodiesel Development and Characterization for Use as a Fuel in Compression Ignition Engines

A. K. Agarwal
Engine Research Center,
University of Wisconsin, Madison
1500 Engineering Drive, Madison, WI 53706
E-mail: avinashagarwal@hotmail.com

L. M. Das
Centre for Energy Studies
Indian Institute of Technology, Delhi
Hauz Khas, New Delhi-110016, India

Neat vegetable oils pose some problems when subjected to prolonged usage in CI engine. These problems are attributed to high viscosity, low volatility and polyunsaturated character of the neat vegetable oils. These problems are reduced to minimum by subjecting the vegetable oils to the process of transesterification. Various properties of the biodiesel thus developed are evaluated and compared in relation to that of conventional diesel oil. These tests for biodiesel and diesel oil include density, viscosity, flash point, aniline point/cetane number, calorific value, etc. The prepared biodiesel was then subjected to performance and emission tests in order to evaluate its actual performance, when used as a diesel engine fuel. The data generated for various concentrations of biodiesel blends were compared with base line data generated for neat diesel oil. It was found that 20 percent blend of biodiesel gave the best performance amongst all blends. It gave net advantage of 2.5 percent in peak thermal efficiency and there was substantial reduction in smoke opacity values. This blend was chosen for long term endurance test. The engine operating on optimum biodiesel blend showed substantially improved behavior. A series of engine tests provided adequate and relevant information that the biodiesel can be used as an alternative, environment friendly fuel in existing diesel engines without substantial hardware modification. [DOI: 10.1115/1.1364522]

Introduction

The world is presently confronted with the twin crises of fossil fuel depletion and environmental degradation. Indiscriminate extraction and lavish consumption of fossil fuels have led to reduction in underground based carbon resources. The search for an alternative fuel, which promises a harmonious correlation with sustainable development, energy conservation, management, efficiency and environmental preservation, has become highly pronounced in the present context. For the developing countries of the world, fuels of bio-origin can provide a feasible solution to the crises. The fuels of bio-origin may be alcohol, vegetable oils, biomass, and biogas. Some of these fuels can be used directly while others need to be formulated to bring the relevant properties close to conventional fuels. The power used in the agricultural and transportation sector is essentially based on diesel fuels and it is, therefore, essential that alternatives to diesel fuels be developed. Given the recent widespread use of diesel fuel in various sectors, this study concentrates on assessing the viability of using vegetable oils in the existing diesel engines. Vegetable oils have comparable energy density, cetane number, heat of vaporization, and stoichiometric air/fuel ratio with mineral diesel fuel. The large molecular sizes of the component triglycerides result in the oils having higher viscosity compared with that of mineral diesel fuel. The viscosity of liquid fuels affects the flow properties of the fuel, such as spray atomization, consequent vaporization, and air/fuel mixing. The problem of viscosity has an adverse effect on the combustion of vegetable oils in the existing diesel engines. Besides some problems crop up in the associated fuel pump and injector system. An acceptable alternative fuel for engines has to fulfill the environmental and energy security needs without sacrificing operating performance [1].

Methanol and ethanol are two abundantly available alternative fuels, which possess the potential to be produced from biomass sources. These fuels can be successfully used as diesel engine fuels by preparing biodiesel. Transesterification process utilizes methanol or ethanol and vegetable oils as the process inputs. This indirect route of utilizing alcohol as a diesel engine fuel is definitely a superior route as the toxic emissions containing aldehydes are drastically reduced. The problem of corrosion of various engine parts utilizing alcohol as fuel is also solved by way of transesterification [2].

The present work is carried out using a typical vegetable oil by formulating its properties closer to the conventional diesel oil. System design approach has taken care to see that these modified fuels can be utilized in the existing diesel engine without any substantial hardware modifications. India is producing around 6.7×10^6 tons of non-edible oils such as linseed, castor, karanji (*Pongamia glabra*), neem (*Azadirachta indica*), palash (*Butea monosperma*), and kusum (*Schlelchera trijuga*). Some of these oils produced even now are not being properly utilized, and it has been estimated that some other plant-based forest derived oils have a much higher production potential. It will be expensive and time consuming to incorporate even a minor design alteration in the system hardware of a large number of existing engines operating in the rural agricultural sector of the country. Keeping this in view, several modes of fuel formulation such as blending, transesterification and emulsification were adopted in the present work to identify the most appropriate mechanism.

Vegetable oils have about 90 percent of the heat content of diesel fuel and they have a favorable output/input ratio of about 2 to 4: 1 for unirrigated crop production. The current prices of vegetable oils in the world are nearly competitive with petroleum fuel price. From amongst the large number of vegetable oils available in the world, if any specific oil needs to be adopted as a continuing energy crop, it is then essential that an oilseed variety having higher productivity and oil content must be produced. Nevertheless, valid technologies must be developed for the use of vegetable oils as an alternative diesel fuel that will permit crop production to proceed in an emergency situation. Vegetable oils in its raw form cannot be used in engines. They have to be converted to a more engine-friendly fuel called biodiesel.

Biodiesel is a chemically modified alternative fuel for diesel

Contributed by the Internal Combustion Engine Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the ICE Division Apr. 24, 2000; final revision received by the ASME Headquarters Dec. 27, 2000. Associate Editor: D. Assanis.

engines, derived from vegetable oil fatty acids, and animal fat. In its simplest form, the carbon cycle of vegetable oils consist of the fixation of carbon and the release of oxygen by plants through the process of photosynthesis and then combining of oxygen and carbon to form CO_2 through processes of combustion or respiration. It is appropriate to mention here that the CO_2 released by petroleum diesel was fixed from the atmosphere during the formative years of the earth, whereas the CO_2 released by biodiesel gets continuously fixed by plants and may be recycled by the next generation of crops. The carbon cycle time for fixation of CO_2 and its release after combustion of biodiesel is quite small as compared (few years) to the cycle time of petroleum oils (few million years). It is well known that petroleum refiners are now facing new sulfur and aromatic compound specifications. Since biodiesel is a fuel made up of esters derived from oils and fats from renewable biological sources, it has been reported to emit far less regulated pollutants than petroleum diesel fuel [3].

Historical Perspective

The inventor of the diesel engine, Rudolf Diesel, in 1885, used vegetable oils (peanut oil) as a diesel fuel for demonstration at the 1900 world exhibition in Paris. Speaking to the Engineering Society of St. Louis, Missouri, in 1912, Diesel said, ''The use of vegetable oils for engine fuels may seem insignificant today, but such oils may become in course of time as important as petroleum and the coal tar products of the present times'' [4]. However, higher viscosity is responsible for various undesirable combustion properties of neat vegetable oils. Four well-known techniques are proposed to reduce the viscosity levels of vegetable oils, namely dilution, pyrolysis, micro-emulsion, transesterification [5]. Transesterification, however, is the current method of choice in the study, which results in a fuel very similar to that of diesel oil. Transesterification was known as early as 1864, when Rochleder described glycerol preparation through ethanolysis of castor oil. Transesterification is a chemical reaction that aims at substituting the glycerol of the glycerides with three molecules of monoalcohols such as methanol thus leading to three molecules of methyl ester of the vegetable oil. The idea of chemically altering vegetable oils was noted even before World War II. Walton wrote in 1938 ''to get the utmost value from vegetable oils as fuels it is academically necessary to split the glycerides and to run on the residual fatty acid'' because ''the glycerides are likely to cause an excess of carbon in comparison'' [6].

The process of utilizing biodiesel in the IC engines for transport as well as other applications, is gaining momentum recently. IEA has recognized biodiesel as an alternative fuel for the transportation sector. The European Commission proposed a 12 percent market share for biofuels by the year 2020 [7]. Kaltschmitt et al. conducted a study, which shows that bioenergy carriers offer some clear ecological advantages over fossil fuels such as conserving fossil energy resources or reducing the greenhouse effect [8].

Knothe investigated the influence of fatty acid structure on the performance as a diesel fuel [9]. Pischinger et al. found monoesters to be technically attractive in terms of good miscibility with diesel oil, almost similar volumetric heat content, adequate viscosity and cetane number [10]. The high carbon residue indicated by the Conradson value and high viscosity are due to the large molecular mass and chemical structure. The high carbon residue is likely to lead to heavy smoke emission from an engine [11].

Hemmerlein et al. established that all of the engines run on rapeseed oil passed the ECE R49 regulation relating to CO, HC, NO_x, and soot emissions in the 13-mode test. the emissions of aromatics, aldehydes, ketones, and particulate matter were higher when rapeseed oil was used. PAH emissions were reduced in larger cylinder IDI engines and elevated in other engine types [12]. Vegetable oil based fuels are biodegradable, non-toxic, and significantly reduce pollution. Reports on the use of biodiesel in diesel engines indicate a substantial reduction in S, CO, PAH, smoke, noise, and particulate emissions [13–15]. Furthermore, contribution of biodiesel to greenhouse effect is insignificant, since CO_2 emitted during combustion is recycled in the photosynthesis process occurring in the plants and is used as a raw material for biodiesel production [14,16]. In 1992, the bi-state development agency (Transit Authority for the Greater St. Louis Metropolitan Area) began a limited 3320 km (2000 mile) test using 30 percent blend of biodiesel and No. 2 diesel fuel in two Ford vans powered by Navistar 7.3 liter engines. They observed that there was no change in the fuel mileage, smoke opacity was reduced by 30 percent, and exhaust odors were less offensive. There were no adverse engine effects. The same study was expanded to include 20 vans and after 41,500 km (25,000 miles), the results were the same [3,17].

In Austria, biodiesel from rapeseed oil is commercialized. The first industrial plant for biodiesel production capacity of 10,000 tons per year in the world went into operation in Austria, in 1991. RME is produced in co-operative and industrial-scale plants. An interesting point is that biodiesel is produced from a mix of RME, SME, and FME (used frying oil methyl ester) in Austria. Tax benefits in Austria and Germany encourage the use of 100 percent biodiesel fuel in ecologically sensitive areas and agricultural and mountainous region [18,19]. In Ireland, two pilot projects on the use of biodiesel (RME) in commercial vehicles are reported. In the first project, pure biodiesel was used in buses, trucks, and in pleasure cruisers. The RME used was obtained through oil extraction followed by esterification on a small-scale tractor-mounted esterification unit. The second project involved a comparative analysis of the behavior of a number of vehicles operated on RME and SME, imported from UK and Italy, respectively. Vehicle testing started in mid-1996. A principal disadvantage of rape oilseed as a source of biodiesel in Ireland was its economic feasibility, as it is up to 0.25 ECU/liter (about 6.6 ECU/GJ) more expensive than its fossil-based equivalent. In Denmark, two companies produced non-food rapeseed oil, on a commercial basis, for export. At present, there is no commercial production of biodiesel in Denmark. During a test, four city buses were running on biodiesel for 3 months. The results were promising but not satisfactory from environmental point of view. The Danish National Transport Plan ''Trafik 2000'' mentions the use of biofuels as one out of five major instruments to be used to reduce CO_2 emissions from the transport sector. However, there is no commitment for large-scale use within the government [20]. In the United States, fuel tax assistance or other governmental aid or regulations could propel biodiesel into a high volume, lower cost production track that enables it to compete head-to-head with diesel fuel for a variety of applications [4].

Vegetable Oil and Diesel Fuel: A Comparison

The ideal diesel fuel molecules are saturated non-branched hydrocarbon molecules with carbon number ranging between 12 to 18 whereas vegetable oil molecules are triglycerides generally with no branched chains of different lengths and different degrees of saturation. It may be noticed that vegetable oils contain a substantial amount of oxygen in their molecular structure. Fuel properties for the combustion analysis of vegetable oils can be grouped conveniently into physical, chemical, and thermal properties. Physical properties include viscosity, density, cloud point, pour point, flash point, boiling range, freezing point, and refractive index. Chemical properties comprise chemical structure, acid value, saponification value, iodine value, peroxide value, hydroxyl value, acetyl value, overall heating value, ash and sulfur contents, sulfur and copper corrosions, water and sediment residues, oxidation resistance, ignitability and thermal degradation products. The thermal properties are distillation temperature, thermal degradation point, carbon residue, specific heating content, and thermal conductivity.

American society of testing methods (ASTM) has prescribed certain tests and their limits for diesel fuel to be used in CI engines. For any alternative fuel to be suitable for long-term engine operation without engine modifications, it should be in conformity or within close range to these ASTM limits [21].

The characteristics of the variety of available vegetable oils fall within a fairly narrow band and are close to those of diesel oil. Vegetable oils have about 10 percent less heating value than diesel oil due to the oxygen content in their molecules. The kinematic viscosity is, however, several times higher than that of diesel oil. The high viscosity, 35–45 cSt as against about 4.0 cSt for diesel oil at 40°C, leads to problems in pumping and atomization in the injection system of a diesel engine. The high carbon residue indicated by the conradson value and high viscosity are due to the large molecular mass and chemical structure. The high carbon residue is responsible for heavy smoke emission from an engine. Vegetable oils have poor volatility characteristics, prohibiting their use in SI engines. Vegetable oils are suitable as fuels only for CI engines. The cetane number of most of the vegetable oils fall in the range 35–40, which is short of the minimum cetane number of 40 required for diesel engine fuel.

Since neat vegetable oils are not suitable as fuels for a diesel engine, they have to be modified to bring their combustion-related properties closer to those of diesel oil. This fuel modification is mainly aimed at reducing the viscosity to get rid of flow related problems. The methods employed include the heating of vegetable oil, thermal cracking of the highly complex vegetable oil molecule into lighter ones, and transesterification, in which methanol/ethanol substitutes glycerol, splitting the vegetable molecule into three lighter molecules. Vegetable oils have larger molecules, up to four times larger than typical diesel fuel molecules. The high molecular weights of vegetable oils result in low volatility as compared to diesel fuel, which leads to the oils sticking to the injector or cylinder walls [4]. Oils then undergo oxidative and thermal polymerization, causing a deposition on the injector, forming a film that continues to trap fuel and interfere with combustion and leads to more deposit formation, carbonization of injector tips, ring sticking, and lubricating oil dilution and degradation. The accompanying inefficient mixing with air contributes to incomplete combustion. The combination of high viscosity and low volatility of vegetable oils causes poor cold engine start-up, misfire, and ignition delay.

A host of plant and forest resources are available in the world from which different vegetable oils can be produced and formulated for use in diesel engines. The present work has been carried with such an objective, where a non-edible oil has been appropriately modified by way of esterification and subsequently used for running the diesel engine.

Linseed oil was selected for the present investigation, as it is available in large quantities and non-edible in nature. Linseed oil is obtained from the dried ripe seed of the flex plant, *Linum usitaissimum*, grown in the temperate areas of the world. Its viscosity is lower than most of other vegetable oils. It has high linoleic acid content. Linoleic acid is a straight chain molecule of 18 carbon atoms [$C_{17}H_{29}COOH$] with three double bonds at 9-10, 12-13, and 15-16 carbons. Its high degree of unsaturation is responsible for the drying properties of the oil. Linseed oil has lower heating value (LHV) of 39.75 MJ/kg (dry). The comparison of the properties of linseed oil and diesel oil is given in Table 1.

Table 1 Comparison of properties of linseed oil with diesel oil [9]

Properties	Diesel oil	Linseed oil
Specific gravity	0.835	0.935
Net calorific value (MJ/Kg)	45.158	39.75
Kinematic viscosity at 37.8°C (cSt)	3.9	29.3
Color	Light brown	Pale yellow
Stoichiometric air fuel ratio	14.9	12.08

Engine System

Nearly all agricultural tractors, pump sets, farm machinery, and transport vehicles have DI diesel engines. With a view of introducing biodiesel to such systems, it is appropriate to mention here that if these engines run over long periods with vegetable oils, or mixtures of diesel fuel with a high content of unmodified vegetable oils, deposits get formed on several engine parts, which subsequently cause problems in fuel spray. Engines do not take up high load level, even after a few hundred hours and at last, the engine fails catastrophically.

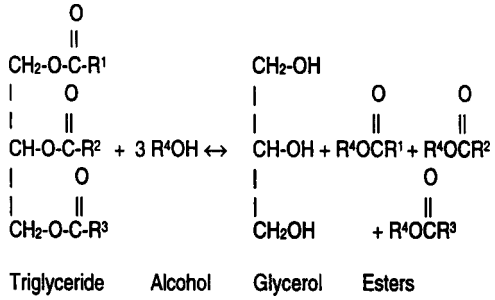
Keeping the specific features of diesel engines in mind, a typical engine system widely used in the agricultural sector in developing countries, has been selected for the present experimental investigation. This is a single cylinder, direct injection, water-cooled, portable diesel engine of 4 kW rating with an alternator coupled to it. The engine was provided with suitable arrangements, which permitted wide variation of controlling parameters. This unit, manufactured by Perry & Co., India, is a compact, portable captive 4 kW gen-set run by diesel fuel. It is widely used mostly for agricultural purposes and in many small and medium scale commercial purposes. This is a single cylinder, four-stroke, vertical, water-cooled system having a bore of 85 mm and a stroke of 110 mm. At the rated speed of 1500 rpm, the engine develops 4 kW power output in pure diesel mode. The engine can be started by hand cranking, using decompression lever. The engine is provided with a centrifugal speed governor. The compression ratio is 16.7. The inlet valve opens 4.5 deg BTDC and closes 35.5 deg ABDC. The exhaust valve opens 35.5 deg BBDC and closes 4.5 deg ATDC. The test engine is directly coupled to a 220 V, single phase ac generator of sufficient capacity to absorb the maximum power produced by the engine.

Esterification Process

The use of biodiesel fuel is an effective way of substituting diesel fuel in the long run. One important conclusion that can be drawn from the work done earlier in our laboratory is that vegetable oils as fuel cannot be used directly in the engine. Several problems crop up if the unmodified fuel is used and viscosity is the major offender. It has been observed that esterification is a very effective way to modify the vegetable oil structure and reduce its viscosity. If vegetable oil processing costs can be reduced through increased plant capacities, it would become an economically viable alternative to diesel fuel.

The formation of methyl esters by transesterification of vegetable oils requires three moles of alcohol stoichiometrically. However, transesterification is an equilibrium reaction in which excess alcohol is required to drive the reaction close to completion. Fortunately, the equilibrium constant favors the methyl esters such that only a 5:1 molar ratio of methanol: triglyceride is sufficient for a 95 percent–98 percent yield of ester. It might be anticipated that in such an equilibrium system, the observed phase-separation of the by-product, glycerol would play a major role in achieving conversions close to 100 percent. Various catalysts were tried for the process of transesterification by several earlier researchers, e.g., magnesium and calcium oxides and carbonates, basic and acidic macro-reticular organic resins, alkaline alumina, phase transfer catalysts, sulfuric acid, P-toluenesulphonic acid, and dehydrating agents as cocatalysts [22]. Catalysts reported to be effective at room temperature were alkoxides and hydroxides. The chemical reaction of the transesterification process is shown

below.



R^1, R^2, R^3 and R^4 represent various alkyl groups [23]. For methanolysis, two distinct phases are present as the solubility of the oil in the methanol is low. The reaction mixture was stirred vigorously. Optimum reaction conditions for the maximum yield of methyl esters were found to be 0.8 percent (based on weight of oil) sodium methoxide catalyst and 67 percent excess methanol at room temperature for 2.5 hours. Glycerol phase separation was very small or did not occur when less than 67 percent of the theoretical amount of methanol was used. There are some alternative methods available for the process of esterification. An alternative reactive substance is sodium hydroxide (NaOH), which is cheaper but it requires a higher reaction temperature (70°C) and produces a useless by-product. As far as KOH is concerned, it reacts at room temperature and the by-product, potassium sulfate, can be sold as a fertilizer. Methanol is then removed by distillation to raise the flash point of the ester produced. All traces of methanol, KOH, free fatty acids (FFA), chlorophyll, etc., go into the glycerin phase. Glycerin is processed in two stages. The first stage glycerin of 90 percent–95 percent purity is further esterified to 98 percent purity in the second stage. The transesterification plant requires 50 kW per ton of ester produced. Energy required in production of glycerin accounts for 60 percent–70 percent of the total energy used.

The molecular weight of linoleic acid ($C_{17}H_{28}COOH$) in linseed oil molecule is 278. This is also confirmed by the CHNO analysis and infrared spectroscopy. The molecular weight of methanol (CH_3OH) is 32. Linseed oil contains 9.40 percent (18:0) saturated linoleic acid, 20.2 percent (18:1) unsaturated linoleic acid, 12.7 percent (18:2) unsaturated linoleic acid, 53.3 percent (18:3) unsaturated linoleic acid. Thus linseed oil contains 95.6 percent total linoleic acid [24]. The density of linseed oil is 0.935 and methanol is 0.80. Every liter of linseed oil requires 643 ml of methanol. For esterification in laboratory, 2 liters of linseed oil was heated up to 70°C in a round bottom flask. 16 g of KOH was dissolved in 1286 ml of methanol in a separate vessel, which was poured into round bottom flask while stirring the mixture continuously using constant stirring mechanism. The mixture was stirred and maintained at 70°C for 1 h and then allowed to settle under gravity in a separating funnel. About 2.3 liter of ester separated from 230 ml glycerol and formed the upper layer in the separating funnel. The separated ester was mixed with 250 ml of hot water and allowed to settle under gravity for 24 h. The catalyst got dissolved in water, which was separated. Moisture was removed from this purified ester using silica gel crystals. The ester was then blended with petroleum diesel oil in various concentrations for preparing biodiesel blends to be used in CI engine for conducting various engine tests. The level of blending, for convenience is referred as Bxx. The xx indicates the amount of biodiesel in percentage in the blend (i.e., a B20 blend is 20 percent biodiesel and 80 percent diesel oil) [25].

Biodiesel Characterization

Several tests were conducted to characterize biodiesel in relation to diesel oil in order to evaluate various physical, chemical, and thermal properties such as density, API gravity, viscosity (40°C and 100°C), flash point, pour point, aniline point, cetane number, calorific value, etc. The results obtained are shown in Table 2.

Table 2 Various characterization properties of biodiesel and diesel

Properties	ASTM Test No.	Linseed oil	Diesel Oil	LOME Biodiesel	20% biodiesel blend
Density	---	0.935	0.855	0.874	0.859
API gravity	---	19.833	33.997	30.399	33.226
Viscosity (40°C)	D445	23.93	3.06	3.59	3.20
Viscosity (100°C)	D445	6.00	1.05	1.16	1.32
Flash point (°C)	D93	186	76	172	128
Pour point (°C)	---	---	-16	-15	-16
Aniline Point (°C)	---	---	69	83	73
Diesel Index	---	---	53.10	55.14	54.29
Cetane number		---	50	52	51
Sulfur (ppm)	D129	---	2500	50	2000
Ash Content (%)	D482	---	0.01	0.002	0.008
Calorific Value (KJ/Kg)	---	---	43,800	40,374	43,200

The process of esterification brings about a drastic change in the density of linseed oil and the LOME; i.e., neat biodiesel has almost similar density to that of diesel. Neat biodiesel was miscible in any proportion with that of mineral diesel oil. The optimized blend of biodiesel has density very close of diesel oil.

Linseed oil has a higher viscosity at 40°C compared to diesel oil. This property of linseed oil makes it unsuitable for direct use in CI engines. By converting linseed oil into its methyl esters, the viscosity comes down closer to diesel oil. The diesel oil viscosity is normally between 3 and 4 cSt at 40°C and it varies depending upon its constituents. Thus fuel oil, which has viscosity within this range, does not pose any handling problems to the fuel handling systems in the existing diesel engines. Hence, no hardware modifications are required for handling this fuel in the existing system. The viscosity at 100°C of biodiesel also closely follows diesel oil. This property makes it suitable for fuel use in the existing engine systems.

The flash point of linseed oil was lowered after transesterification but it was still higher than diesel oil. Fuels with flash point above 66°C are regarded as safe. Thus, biodiesel is an extremely safe fuel to handle compared to diesel oil. Even 20 percent biodiesel blend has a flash point much above that of diesel oil, making it a preferable choice as far as safety is concerned. Pour point is a criterion used for low temperature performance of the fuels. Pour point analysis results suggest that the performance of biodiesel is as good as diesel oil in cold surroundings.

Cetane number of biodiesel fuel was evaluated using an alternative chemical method. This method (aniline point method) was first tested for determining the cetane number of diesel fuel, whose cetane number was already known. The cetane number evaluated by this method and the cetane number given by the fuel supplier matched and this established the precision of the method. Then this method was applied to determine cetane number of biodiesel fuel. In this method, aniline is used as a reagent to evaluate aniline points for these fuels. Diesel index is calculated using aniline point and API gravity. The calculated values of diesel index are given in the table. The diesel index is three units higher than cetane number. These results reveal that biodiesel has a higher cetane number than petroleum diesel oil. Even 20 percent blend of biodiesel showed improvement in cetane number. Thus, biodiesel is an attractive fuel by combustion and knocking point of view and it can be used as a cetane point improver. Even lower concentrations in the range of 3 percent–5 percent of biodiesel blended with diesel are found to improve cetane number of diesel oil by various researchers.

Benzoic acid was used as a reference fuel to determine the calorific value of biodiesel and linseed oil. The results of the tests

conducted on various fuels using Bomb calorimeter are given in Table 2. The residue left after combustion of diesel oil in the crucible was in the form of black carbon ash pellets (charcoal like foamy substance). The quantity of such a foamy charcoal like ash substance was very low in case of LOME combustion. This typical observation points out lesser ash content and cleaner combustion characteristics of LOME. This fact was subsequently supported by the engine emission tests.

Engine Tests

A series of exhaustive engine tests were carried out on CI engine using diesel and biodiesel blends separately as fuels at 1500 RPM. The experimental data generated were calculated, presented through appropriate graphs. Performance and emission test was conducted on various biodiesel blends in order to optimize the blend concentration for long term usage in CI engines. This test was carried out on an engine, which has already been subjected to preliminary run-in. This test is aimed at optimizing the concentration of ester in the biodiesel blends to be used for long term engine operation. To achieve this, several blends of varying concentrations were prepared ranging from 0 percent (Neat diesel oil) to 100 percent (Neat biodiesel) through 5 percent, 10 percent, 15 percent, 20 percent, 25 percent, 30 percent, 40 percent, 50 percent, 75 percent. These blends were then subjected to performance and emission tests on the engine. The performance data was then analyzed from the graphs recording power output, torque, specific fuel consumption, and smoke density for all the blends of biodiesel. The optimum blend was found out from the graphs based on maximum thermal efficiency and smoke opacity considerations.

The major pollutants appearing in the exhaust of a diesel engine are smoke and the oxides of nitrogen. For measuring the smoke opacity, Hartridge Exhaust smoke-meter (model No. APM 700 M), was used. Hartridge smoke meter consists of two identical tubes, a smoke type and a clean air tube. A pressure relieve valve allows a regulated quantity of exhaust through the smoke tube. During smoke density measurements, a light source (45-W bulb) at one end of the smoke tube projects a light beam through smoke, which at the other end falls on a photoelectric cell. Clean air tube is used for initial zero setting. Of the light beam projected across a flowing stream of exhaust gases, a certain portion of light is absorbed or scattered by the suspended soot particles in the exhaust. The remaining portion of the light falls on a photocell, generating a photoelectric current, which is a measure of smoke density. A micro-voltmeter is connected to the photoelectric cell with its scale graduated 0–100, indicating the light absorbed in Hartridge Smoke Meter unit. Zero reading corresponds to no smoke (clean air), whereas 100 reading refers to dense smoke, which allows no light to pass through [26].

The brake specific fuel consumption is not a very reliable parameter to compare the two fuels as the calorific value and the density of the blend follow a slightly different trend. Hence brake specific energy consumption is a more reliable parameter for comparison. However this parameter is also compared in the present study to compare volumetric consumption of the two fuels. For an optimum biodiesel system, the blend concentration has been determined based on the following performance and emission-related parameters.

Maximum thermal efficiency at all loads
Minimum brake specific energy consumption
Better smoke opacity values
Lower frictional horsepower

Based on these parameters, the curves were compared to base-line diesel curve in two separate groups in order to optimize blend concentration. In the first group, lower concentrations up to 25 percent were compared with that of diesel oil and in the second group, higher concentrations from 30 percent to 100 percent were compared to that of base-line data.

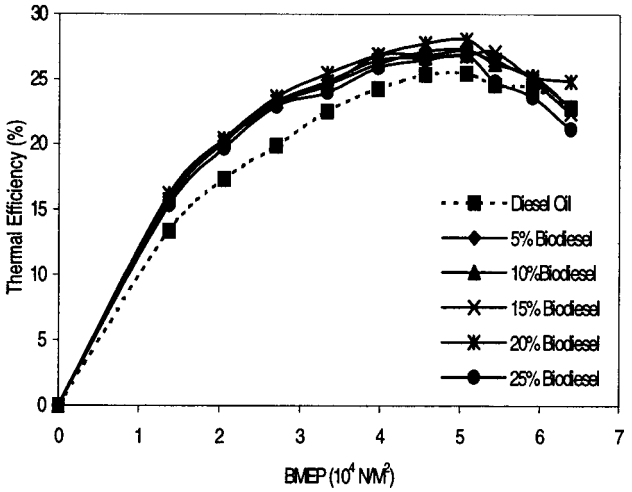


Fig. 1 Comparison of thermal efficiency vs. BMEP curves for lower concentrations of biodiesel

Comparison for Lower Blend Concentrations

The various characteristic curves for lower concentrations of biodiesel blend are compared in Figs. 1–4.

Comparison for Higher Concentration Blends

The various characteristic curves for higher concentrations of biodiesel blend are compared in Figs. 5–8.

The trend of the thermal efficiency curves (Fig. 1 and Fig. 5) has generally improved by mixing biodiesel in mineral diesel oil. The thermal efficiency of the engine is found to improve by increasing concentration of biodiesel in the blend. The possible reasons may be more complete combustion, and additional lubricity of biodiesel hence, with increasing amount of biodiesel, the frictional horsepower losses in the engine follow a declining trend. Additional experiments were conducted to determine the additional lubricity aspect of biodiesel fuels. A series of experiments were conducted to compare lubricity of various concentrations of LOME in biodiesel blends. These long duration non-engine tests were conducted employing reciprocating motion in SRV optimol wear tester in order to evaluate the coefficient of friction, specific

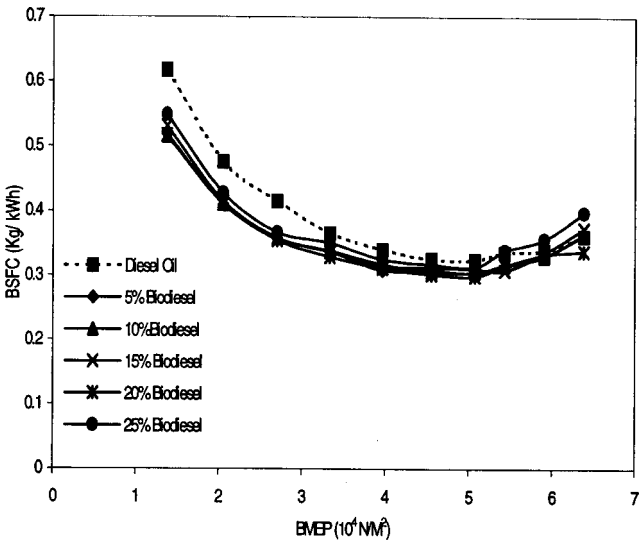


Fig. 2 Comparison of BSFC vs. BMEP curves for lower concentrations of biodiesel blend

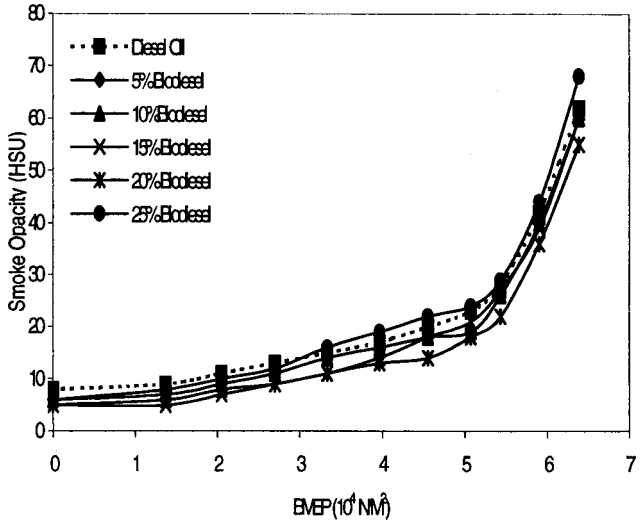


Fig. 3 Comparison of smoke opacity vs. BMEP curves for lower concentrations of biodiesel blend

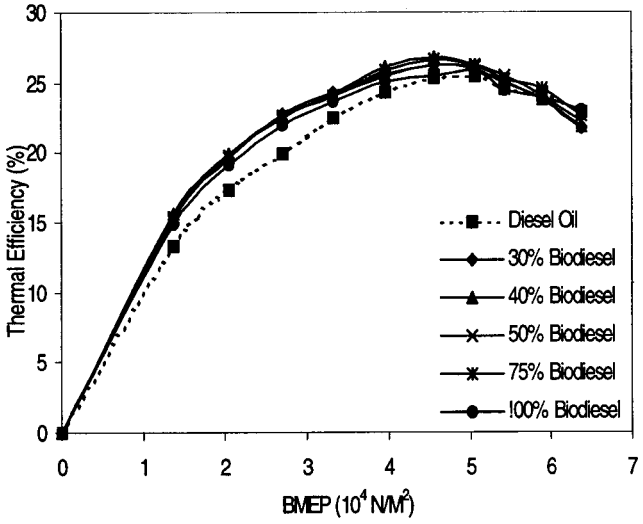


Fig. 5 Comparison of thermal efficiency vs. BMEP curves for higher concentrations of biodiesel blend

wear rates etc. The extent of the damage, coefficient of friction and specific wear rates were found to decrease with increase in percentages of LOME in the biodiesel blend [27]. These tests on SRV optimol wear tester for quantification of lubricity reflected that the biodiesel lubricating fluids show lower coefficient of friction hence lower frictional losses in diesel engines.

This reduction in frictional losses is also reflected by the decreasing trend followed by unaccounted losses, which can be noticed from the heat balance sheets for various concentrations. The energy saved by decrease in frictional horse-power makes additional contributions towards useful energy, cooling losses, and exhaust losses. This reflects in increased thermal efficiency, cooling losses and exhaust losses. It is evident from the graphs of exhaust temperature vs. BMEP (Fig. 4 and Fig. 8) that the exhaust gas temperature is also increasing along with thermal efficiency as a function of blend concentration initially. Smoke opacity for biodiesel blends (Fig. 3 and Fig. 7) is also noticed to be generally lower than that of diesel oil. It also repeats the trend followed by thermal efficiency curves. There are two main reasons: the higher thermal efficiency means, better and complete combustion and lesser amount of unburned hydrocarbons in the engine exhaust thus improving smoke opacity values.

The molecule of biodiesel, i.e., LOME, contains some amount of oxygen that takes part in combustion and this may be a possible reason for more complete combustion. The oxygen molecule present in biodiesel molecular structure may be readily available for oxygen. However, it was noticed that after a certain limit of biodiesel concentration, the thermal efficiency trend is reversed and it starts decreasing. This behavior of biodiesel fuel needs advanced investigations.

An important observation is that all the blends have a higher thermal efficiency than the baseline data of diesel fuel. A graph between the concentration of ester blend and improvement in peak thermal efficiency for various concentrations of biodiesel blend is plotted in Fig. 9.

This graph reflects that ester blend with concentration of 20 percent gave maximum improvement in peak thermal efficiency. These observations were taken six times and there was hardly any difference in the values of the performance and emission parameters calculated. However, in order to minimize any experimental inaccuracies that might have got in, the average of the six readings

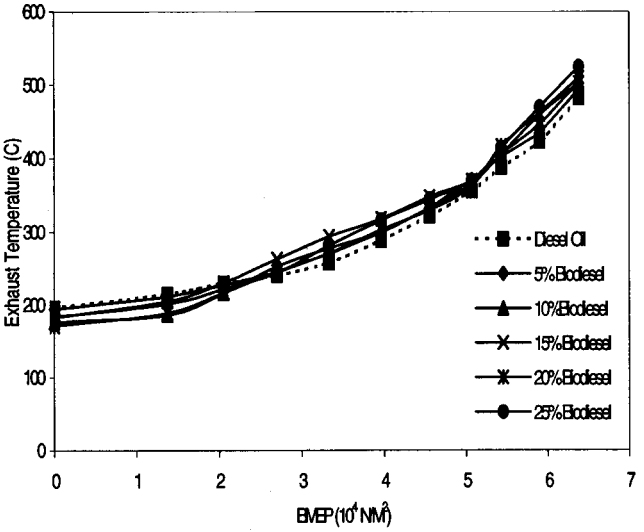


Fig. 4 Comparison of exhaust temperatures vs. BMEP curves for lower concentrations of biodiesel

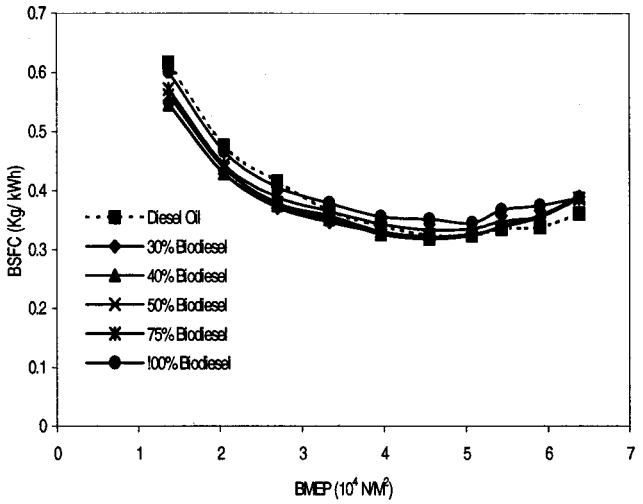


Fig. 6 Comparison of BSFC vs. BMEP curves for higher concentration of biodiesel blend

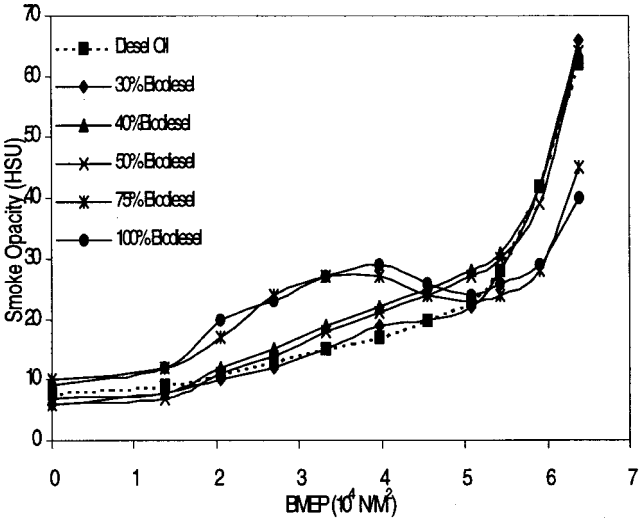


Fig. 7 Comparison of smoke opacity vs. BMEP curves for higher concentrations of biodiesel blend

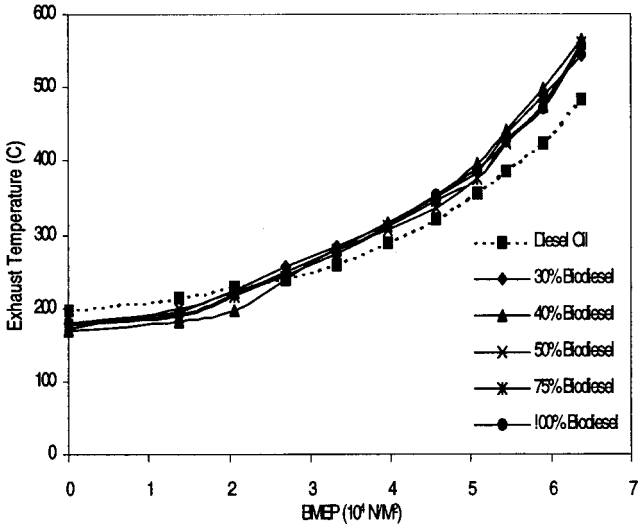


Fig. 8 Comparison of smoke temperature vs. BMEP curves for higher concentrations of biodiesel blend

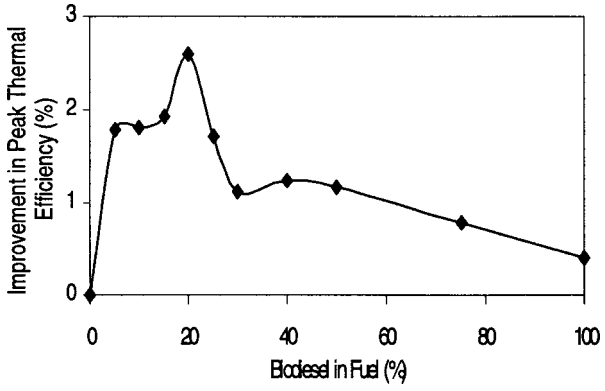


Fig. 9 Improvement in peak thermal efficiency vs. concentration of biodiesel blend curve

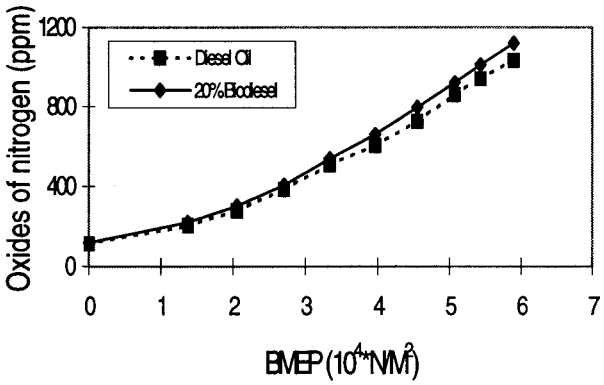


Fig. 10 Concentration of oxides of nitrogen vs. BMEP

were used for calculations. This blend of 20 percent also gave minimum brake specific energy consumption and minimum smoke opacity. Hence this blend was selected as the optimum blend for further investigations and long-term engine operations. Long-term endurance test was conducted with 20 percent biodiesel blend and diesel oil on two similar engines. For diesel fuel, the most important pollutants are smoke and NO_x hence the optimum biodiesel concentration is also subjected to evaluation of NO_x emissions.

From the NO_x curves given in Fig. 10, two important observations were made. First, NO_x emissions are a direct function of engine loading. This is expected because with increasing load, the temperature of the combustion chamber increases and NO_x formation is a strongly temperature dependent phenomenon. Second important observation is that the NO_x emissions in the case of biodiesel fuel are higher by approximately 5 percent. These higher NO_x emissions may be due to higher temperatures of the combustion chamber using biodiesel. This is also evident from higher exhaust temperatures from the biodiesel-fueled engine.

The NO_x emissions from B20 fueled diesel engines were found to be 2 percent higher than conventional diesel oil according to a report on biodiesel from National Biodiesel Board of US [1]. In our study, 20 percent biodiesel blend gave 5 percent higher NO_x emissions. This difference may be because of difference in engine geometry and compression ratio. The B20 is also reported to provide additional lubricity compared to conventional fuels and demonstrate similar fuel consumption, horsepower and torque, as petroleum fuel [1]. The present study also demonstrated similar results and is in full agreement with the findings of these researchers.

The long-term endurance test for constant speed engines was performed after the performance tests were carried out as specified in the Indian Standard Code ''IS: 10000, part VIII-1980,'' ''Methods of tests for internal combustion engines: Part VIII Performance tests.'' After the completion of performance tests, two similar engines were operated using two different fuels for 32 cycles (each of 16 h continuous running) at rated speed. The test cycle followed is specified in the Table 3 [25].

At the end of each 16-h cycle, both the engines were stopped for necessary servicing, and minor adjustments were carried out in accordance with the manufacturer's schedule, e.g., tappet settings,

Table 3 Test cycle for long-term endurance test

Load (% of rated load)	Running time (hours)
100	4 (including warm-up period for 0.5 hours)
50	4
110	1
No load (idling)	0.5
100	3
50	3.5

and make up oil addition. The amounts of make up oil used during the test were used to establish the lubricating oil consumption rate. The lubricating oil samples were collected from the engines after every 128 h for conducting various tribological studies.

The long-term endurance test suggested that biodiesel fuel does not have any adverse effect on wear of various vital moving parts of the engine. The problems which were present in long term operation of using unmodified vegetable oils as fuel such as injector coking, lube oil dilution, carbon deposits on piston, ring sticking, fuel pump failure, catastrophic failure of the engine were found to have been completely eliminated.

The pure LOME cost 50 percent higher than diesel oil in India hence 20 percent biodiesel blend costs about 10 percent more expensive. This work is carried out only on linseed oil but this work can be extended to a host of vegetable oil available from agricultural and forest resources.

Conclusions

Based on the exhaustive engine tests, it can be concluded that biodiesel can be adopted as an alternative fuel for the existing conventional diesel engines without any major modifications required in the system hardware.

Esterification is a process, which brings about a change in the molecular structure of the vegetable oil molecules, thus bringing down the levels of viscosity and unsaturation of vegetable oils. The viscosity of vegetable oil reduces substantially after esterification. The density and viscosity of the linseed oil methyl ester formed after esterification were found to be very close to petroleum diesel oil. The flash point of LOME was higher than that of diesel oil. The 20 percent biodiesel blend also demonstrated comparatively higher flash point than petroleum diesel oil and was in range of 'safe fuel'. The cetane number for the neat diesel oil obtained by this method is 50 and cetane number of LOME was 52. A 20 percent blending of LOME with diesel oil improved the cetane number of diesel oil. Lower concentrations of LOME in biodiesel blends can be used as a cetane improver. The calorific value of LOME was found to be slightly lower than petroleum diesel oil. All these tests for characterization of biodiesel oil demonstrated that almost all the important properties of biodiesel are in very close agreement with the diesel oil making it a potential candidate for the application in compression ignition engines for partial replacement of diesel fuel.

A diesel engine can perform satisfactorily on biodiesel fuel without any engine hardware modifications. Long-term endurance test using biodiesel proved that biodiesel can be used for substituting mineral diesel oil in long run. The 20 percent biodiesel blend was found to be the optimum concentration for biodiesel blend, which improved the thermal efficiency of the engine by 2.5 percent, reduced the exhaust emissions and the brake specific energy consumption. Smoke emissions also reduced appreciably. The higher the concentration of biodiesel blend, higher was the reduction in exhaust smoke levels. Exhaust temperatures increased as a function of the concentration of biodiesel blend, i.e., higher the percentage of LOME; higher were the exhaust temperatures. Increase in the exhaust temperature of a biodiesel-fueled engine led to approximately 5 percent increase in NO_x emissions for 20 percent biodiesel blend. This is so because NO_x formation is a highly temperature dependent phenomenon.

Esterification has been found to be an effective technique to prevent some long-term problems associated with utilization of vegetable oils such as fuel filter plugging, injector coking, formation of carbon deposits in combustion chamber, ring sticking, and contamination of lubricating oils. The carbon deposits on piston top and injector coking substantially reduced in biodiesel-fueled system. The performance of biodiesel-fueled engine was marginally better than the diesel-fueled engine in terms of thermal efficiency, brake specific energy consumption, smoke opacity, wear of vital components, and exhaust emissions for entire range of operations. It was conclusively proved that self-lubricity of LOME in biodiesel played a key role in engine performance. Biodiesel is proved to be a potential candidate for partial substitute of mineral diesel oil.

References

[1] ''Biodiesel,'' National Biodiesel Board Report, Jefferson City, MO 65110-4898, 1999.

[2] Scholl, K. W., and Sorenson, S. C., 1993, ''Combustion of Soybean Oil Methyl Ester in a Direct Injection Diesel Engine,'' *SAE paper 930934*, pp. 1450–1461.

[3] Ali, Y., and Hanna, M. A., 1994, ''Alternative Diesel Fuels from Vegetable Oils,'' Bioresource Technology, **50**, pp. 153–163.

[4] Krawczyk, T., 1996, ''Biodiesel: alternative fuel makes inroads but hurdles remain,'' Inform, **7**, pp. 801–814.

[5] Knothe, G., Dunn, R. O., and Bagby, M. O., 1996, ''Technical Aspects of Biodiesel Standards,'' Inform, **7**, pp. 827–829.

[6] Van Dyne, D. L., Weber, J. A., and Braschler, C. H., 1996, ''Macroeconomic Effects of a Community Based Biodiesel Production System,'' Bioresource Technology, **56**, pp. 1–6.

[7] Korbitz, W., 1999, ''Biodiesel Production in Europe and North America, an Encouraging Prospect,'' Renewable Energy, **16**, pp. 1078–1083.

[8] Kaltschmitt, M., Reinhardt, G. A., and Stelzer, T., 1997, ''Life Cycle Analysis of Biofuels under Different Environmental Aspects,'' Biomass Bioenergy, **12**, pp. 121–134.

[9] Agarwal, A. K., 1992, ''Performance Evaluation and Emission Characteristics of a Compression Ignition Engine Using Esterified Biodiesel,'' MTech thesis, *Centre for Energy Studies*, Indian Institute of Technology, Delhi, India, p. 82.

[10] Pischinger, G., Siekmann, R. W., Falcon, A. M., and Fernandes, F. R., 1982, ''Results of Engine and Vehicle Tests with Methyl Esters of Plant Oils as Alternative Diesel Fuels,'' Proceedings of Fifth International conference on ''*Alcohol fuel technology*,'' Auckland, New Zealand.

[11] Rao, P. S., and Gopalkrishnan, K. V., 1991, ''Vegetable Oils and Their Methyl Esters as Fuels for Diesel Engines,'' Indian J. Technol., **29**, pp. 292–297.

[12] Hemmerlein, N., Korte, V., Richter, H., and Schoroder, G., 1991, ''Performance, Exhaust Emissions and Durability of Modern Diesel Engines Running on Rapeseed Oil,'' *SAE Technical paper 910848*, SAE, Warrendale, PA.

[13] Murayama, T., 1994, ''Evaluating Vegetable Oils as a Diesel Fuel,'' Inform, pp. 1138–1145.

[14] Bona, S., Mosca, G., and Vamerli, T., 1999, ''Oil Crops for Biodiesel Production in Italy,'' Renewable Energy, **16**, pp. 1053–1056.

[15] Vicente, G., Coteron, A., Matinez, M., and Aracil, J., 1998, ''Application of Factorial Design of Experiments and Response Surface Methodology to Optimise Biodiesel Production,'' Industrial Crops and Products, **8**, pp. 29–35.

[16] Narayah, R., 1992, ''Biomass (Renewable) Resources for Production of Material, Chemicals and Fuels,'' *ACS Symp. Ser. 476*, pp. 1–10.

[17] Anon, 1998, ''The News Source for the Biodiesel Industry,'' *Biodiesel Alert*, 1 (5). National Soydiesel Development Board, American Biofuels Association, Arlington, VA.

[18] Raneses, A. R., Glaser, L. K., Price, J. M., and Duffield, J. A., 1999, ''Potential Biodiesel Markets and their Economic Effect on the Agricultural Sector of the United States,'' Industrial Crops and Products, **6**, pp. 151–162.

[19] Mittelbach, M., ''A European Perspective on Quality Biodiesel,'' in Peterson, C L (Editor), *Commercialization of Biodiesel: producing a quality fuel*, Conference Proceedings, 9–10 July 1997, Boise, ID, University of Idaho, Moscow, ID, pp. 125–131, 1998.

[20] Venendaal, R., Jorgensen, U., and Foster, C. A., 1997, ''European Energy Crops: a synthesis,'' Biomass Bioenergy, **13**, pp. 147–185.

[21] Rakapoulos, C. D., 1992, ''Olive Oil as a Fuel Supplement in DI and IDI Diesel Engines,'' Energy (Oxford), **17**, pp. 787–790.

[22] Nye, M. J., and Southwell, P. H., 1983, ''Esters from Rapeseed Oil as Diesel Fuel,'' Proceedings of ''*Vegetable oil as diesel fuel- Seminar III*,'' Peoria, IL.

[23] Agarwal, A. K., 1998, ''Vegetable Oils versus Diesel Fuel: development and use of biodiesel in a compression ignition engine,'' *TERI Information Digest on Energy (TIDE)*, **8**, pp. 191–203.

[24] Agarwal, A. K., 1996, ''Vegetable Oil Test Fuels for Diesel Engines: formulation and analysis,'' M. Tech minor project, *Center for Energy Studies*, Indian Institute of Technology, Delhi, India, p. 28.

[25] Agarwal, A. K., 1999, ''Performance evaluation and tribological studies on a biodiesel-fuelled compression ignition engine,'' Ph.D. thesis, *Center for Energy Studies*, Indian Institute of Technology, Delhi, India, p. 344.

[26] ''*APM 700 Hartridge Smoke Meter User's Manual*,'' Environtech Instruments Pvt. Ltd. New Delhi, India, 1996.

[27] Agarwal, A. K., Bijwe, J., and Das, L. M., 2001, ''Wear Assessment in a Biodiesel Fueled Compression Ignition Engine,'' ASME-ICED Spring Conference 2001, to be published.

Challenges Involved in Piston Top Ring Designs for Modern SI Engines

R. Rabuté
Perfect Circle Division,
Dana Corporation,
Poissy, France
e-mail: remi.rabute@dana.com

T. Tian
Sloan Automotive Laboratory,
Massachusetts Institute of Technology,
Cambridge, MA 02139
e-mail: tian.tian@mit.edu

Driven by energy conservation and environment protection, modern SI engines are required to have higher and higher power density. Consequently, engines and engine components are becoming tighter and lighter, and engines are more often operated at elevated speed and component temperature. Piston and piston ring designs thus face constant challenge to provide proper control of blow-by, oil consumption, friction, wear, and oil consumption. This paper is intended to give an overview of the problems associated with top ring performance in modern SI engines, the mechanisms behind these problems, and possible solutions. The analysis is based on engine test data and computer models. Major topics covered in this paper include (1) top ring flutter and collapse, and their influence on blow-by; (2) top ring reverse flutter and its influence on oil consumption; (3) contact between top ring and its groove, and its potential influence on ring/groove wear and micro-welding; (4) top ring/liner lubrication and its influence on ring face friction, wear, and scuffing. [DOI: 10.1115/1.1364520]

Introduction

The development of SI engines is characterized by increase of power density accompanied by reduction in fuel consumption and improvement in performance. Additionally, all the new engine development should meet stringent emission regulations and satisfy customer's need for less frequent service. Of all the other important components, piston rings have a major influence on engine friction power losses, oil consumption, blow-by, and wear. Friction from piston rings contributes to 20 percent–30 percent of the entire engine mechanical loss [1]. A significant portion of engine oil consumption is from piston ring pack region [2] and reducing engine oil consumption is important for a couple of reasons. First, oil service intervals need to be increased to reduce the cost of ownership. Second, reducing oil consumption results in less contamination of catalysts by lubricating oil. Blow-by of combustion gases is a major cause of degradation of the lubricating oil. Additionally, high blow-by can reduce power output and increase oil consumption through a PCV system. At last, avoiding top ring scuffing and micro-welding is a constant challenge for piston and piston ring designs.

Certainly, piston and piston ring pack have to be treated as a system to optimize their performance. However, the top ring, as it is the first barrier to blow-by gas and the last barrier for oil flowing into combustion chamber, has significant impact on engine blow-by and oil consumption. Furthermore, severe lubrication condition occurs between the top ring and its groove due to high temperature and high gas pressure as well as between the top ring and the liner when the top ring travels in the upper portion of the liner region due to high temperature, high pressure, and lack of direct oil supply. Therefore, top ring design is also critical for wear and friction.

In this paper, the main features in modern SI engines that have impact on top ring behavior are summarized first. Then, a practical example is given, in which high blow-by and high oil consumption existed over all steady state conditions as well as fleet test with the first proposed piston and piston ring pack. Computer models were then used to analyze the cause for high blow-by at critical conditions, namely, high-speed/no-load and high-speed/WOT conditions. After identifying the key parameters that contribute to high blow-by, modifications on piston and ring pack were proposed and tests with new designs show simultaneous decrease on blow-by and oil consumption at both steady state and real driving cycle conditions. Another example is used to illustrate gas flow back to the combustion chamber through the top ring groove during top ring flutter and the effects of this phenomenon on oil consumption. Finally, the paper shows a study in another engine where severe lubrication conditions existed between the top ring and the groove as well as between the top ring and the liner. Several top ring and groove geometrical parameters were studied using the models to demonstrate how the lubrication conditions can be improved by changing mechanical behaviors between the top ring and its mating parts.

Contributed by the Internal Combustion Engine Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the ICE Division June 16, 2000; final revision received by the ASME Headquarters Dec. 6, 2000. Associate Editor: D. Assanis.

Main SI Engine Evolution Affecting Top Ring Behavior

Top ring behavior can be critically affected by some aspects of new SI engine designs and operating. Before going further, it is useful to list main SI engine evolutions that lead to those top ring aspects to be discussed in this paper.

Higher Engine Speed. Over recent years, we have observed a continuous increase of engine speed in SI engines. For instance, now in Europe, a maximum engine speed of at least 6500 rpm has become common. This evolution, intended to boost the output of the engine, requires several modifications in the piston system. For piston designs, the major impacts of increased engine speed are the problems associated with the kinematics and the resulting NVH. On the piston ring side, increasing engine speed results in more complex ring dynamic behavior and gas flow related to ring dynamics because the ring axial inertia force (Fig. 1) is able to compete with the gas pressures over a greater portion of a cycle. One way to reduce top ring axial inertia force and suppress top ring instability and high blow-by is decreasing top ring axial height, which also serves the purpose of reducing piston crevice volume. Now, the top ring with 1.2 mm axial width becomes standard.

Lightweight Piston Design. To decrease the noise generated by the piston and decrease the engine friction, a reduction of the oscillating masses became necessary, which is mainly a task of piston designs [3]. As a consequence of piston weight reduction,

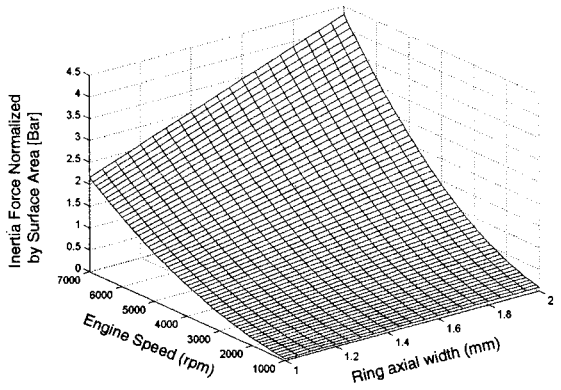


Fig. 1 Maximum ring inertia force due to piston acceleration, normalized by the area of ring flank

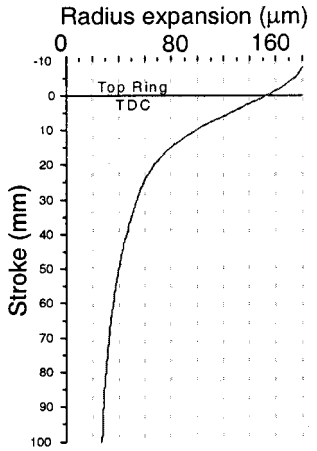


Fig. 3 Bore expansion under WOT

the ring axial width as well as the piston land width have to be decreased. Smaller volumes between the rings leave less possibility to modify the inter-ring pressures. Small areas on the piston lands make it more difficult to have any machining done on the piston.

Aluminum Engine Block Generalization. As the engine block is a heavy individual component, substituting aluminum for cast iron results in weight reduction up to 40 percent–60 percent [4]. Additionally, an aluminum block, because of its higher thermal conductivity, requires less cooling and shortens the engine warm-up phase and consequently reduces startup emissions. The drawbacks of aluminum blocks compared with the cast iron ones are higher thermal expansion of the block and lower block stiffness. In addition, aluminum blocks are usually open deck design. As a result, aluminum blocks bring in some challenges to ring designs:

(i) improving bore/liner material and coating,
(ii) improving oil and gas sealing due to higher cylinder bore distortion and expansion.

In spite of the improvement made by using MLS (Multi Layer Steel) cylinder head gasket and the design change on the cylinder head bolts, the bore distortion and expansion for aluminum block engines are usually still significant (Figs. 2 and 3). Figure 2 shows the arrangement of the bolts on an engine head, which is the major source for fourth order bore distortion. Figure 3 shows the calculated radius expansion of a 2.0 L SI engine with an aluminum block under a WOT condition. As a result of bore expansion, actual ring gap variation during engine operation can be significantly greater than designed cold ring gaps, traditional way of controlling blow-by by changing cold ring gaps becomes less relevant. Furthermore, higher order bore distortion brings problems in ring conformability and oil consumption.

Reducing Fuel Consumption. Much attention has been given to reducing friction power loss in the piston system for improving engine fuel efficiency. These efforts include adding piston skirt coating, improving piston ring coating, reducing total ring tangential loads, and using low friction engine oils. Of all the other improvements, lowering oil viscosity and improving friction modifiers [5] are two major focal points for designing low friction engine oils. A lower oil viscosity combined with an increased of bore wall temperature and a decrease of the oil retention capability of the bore surface topography results in less oil film thickness to lubricate the rings so that rings experience more mixed and boundary lubrication with the liner. Therefore, successfully implementing low friction engine oils must be accompanied with other efforts in piston, ring, and liner designs. Particularly, in order to achieve minimal frictional power loss under all lubrication regimes, the design of the top ring must be able to effectively take advantage of new lubricants because it encounters most severe lubrication condition in the piston system.

Reducing Hydrocarbon Emissions. Unburned hydrocarbons from the exhaust pipe of today's car have to meet the strong emission regulation. These HC emissions result from several different mechanisms. Each mechanism is a path by which some of the fuel escapes from normal combustion. The mechanism that the top ring region is involved is the accumulation of unburned hydrocarbons in the piston crevice volumes (Fig. 4). The relative importance of various sources of HC in terms of their contribution to engine-out HC emissions at part load in a warm engine was investigated [6] and it was found that the piston crevice was responsible for 38 percent of the total HC emissions. As the largest crevice, the piston top land volume, namely, axial heights of crown land and rings, has to be reduced substantially. On the other hand, reducing piston top land height in conjunction with

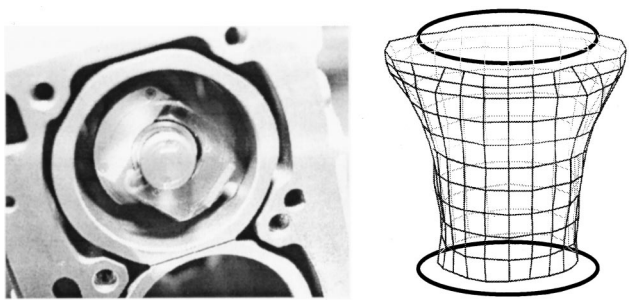


Fig. 2 Static bore distortion

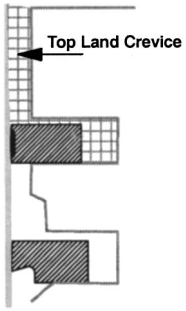


Fig. 4 Top land crevice volume

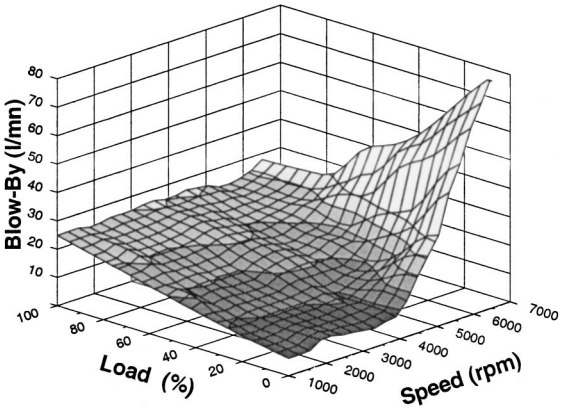


Fig. 5 Engine blow-by map with the original design

the general increase in engine performance led to elevated top ring groove temperature, which becomes as high as 260°C in some extreme applications. This level of temperature can create piston groove damage and microwelding. In practice, additional to new piston/top ring material and coating, a special attention should be given to the contact condition between the top ring and its groove.

In summary, these main engine evolutions change significantly the top ring environment and create some new challenge for the top ring:

(i) top ring instability, namely, top ring radial collapse, flutter, and top ring reverse flutter,
(ii) blow-by control in the engines with aluminum blocks,
(iii) top ring severe lubrication between the ring running surface and the bore, as well as between the flanks of the top ring and the piston groove.

Top Ring Instability and Blow-By Control

The first example is based on an application in a 2.0 L four-cylinder SI engine with an aluminum block. The initial piston and ring pack proposal produced high blow-by at steady state conditions and high oil consumption at both endurance and sport driving test conditions that is primarily consisted with high engine speed and fast transient cycles (Figs. 5 and 6). Furthermore, one can see that the oil consumption during the sport driving condition is much higher than during the endurance test. It was deemed that overall high blow-by and particularly extremely high blow-by at high speed/light load conditions are the cause of high oil consumption under the sport driving condition. As one knows, blow-by gas carries certain amount of oil through piston ring pack and crankcase. In this case, the oil separator of the PCV (positive

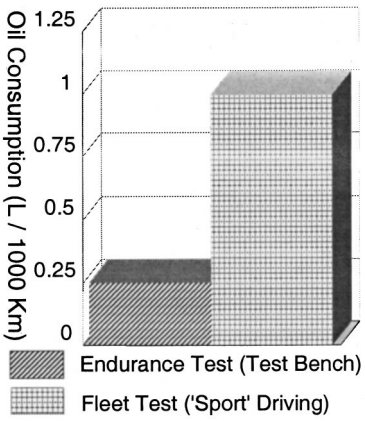


Fig. 6 Oil consumption with the original design

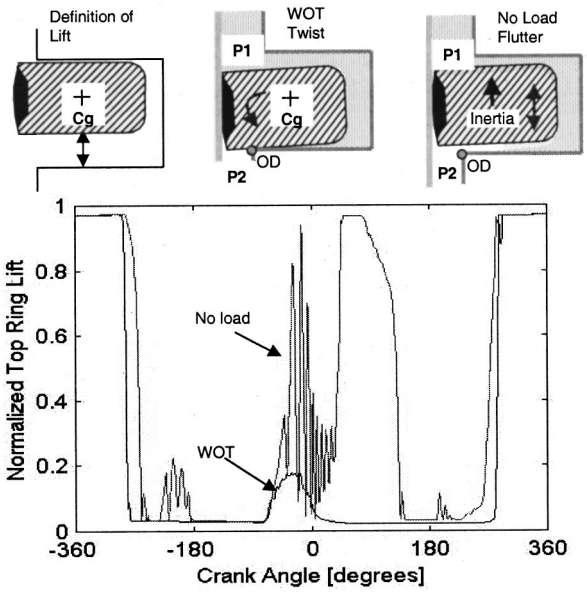


Fig. 7 Influence of engine load on top ring lift at 6000 rpm

crankcase ventilation) system was not able to function properly with excessive blow-by gas so that large amount of oil stays with the recycled blow-by gas and is returned to the combustion chamber through the intake valves, resulting in high oil consumption. Thus, the task was to reduce blow-by under all the operating conditions and consequently the oil consumption.

Top Ring Flutter and Induced Second Ring Radial Collapse Under High-Speed/No-Load Conditions. It is well known that at high-speed/light-load conditions, SI engines often produce much higher blow-by values than WOT conditions. The consensus on the cause is that under high-speed/light-load operating conditions, the top ring loses its ability of sealing gases through the ring/groove or ring/liner interfaces: namely, top ring flutters inside the groove or radial collapse occur. To have more detailed quantitative analysis, the models developed in the MIT Lubrication Consortium were used [7,8]. The ring pack model has been much improved since the first publication [7,8] and one of its new features is that the ring radial collapse was taken into account.

As noticed from Fig. 5, at 6000 rpm, the level of blow-by at the no load condition is much higher than at WOT. It is thus interesting to compare the differences between these two extreme load conditions at 6000 rpm in top ring dynamics (Fig. 7) and land pressures (Figs. 8 and 9).

Figure 7 shows the calculated normalized lift of the top ring at its CG location under no load and WOT conditions with engine speed of 6000 rpm. One can clearly see that under the no load condition the top ring flutters around the TDC of compression/expansion strokes. During top ring flutter, gas flows through the top ring groove all over the circumference and second land pressure and overall gas flow rate through the top ring are dramatically increased (Figs. 9 and 10). Figure 10 shows the instantaneous and average gas flow rate via different paths through the top ring. Hereafter, the sign convention for the gas flow is that positive is flowing down and negative is flowing up. Notice that for the averaged values, the mass flow rate is converted to volumetric flow rate based on the air density at the ambient condition for easier comparison with blow-by values. One can see from the averaged values in Fig. 10 that under the condition of 6000 rpm/no-load, almost all the blow-by to the crankcase is from the gas flow through the top ring groove. The mechanism of ring flutter was described in [7,8] and is not going to be explained here in great details. In short, at high-speed/light load conditions, high upward ring axial inertia force and low cylinder pressure are com-

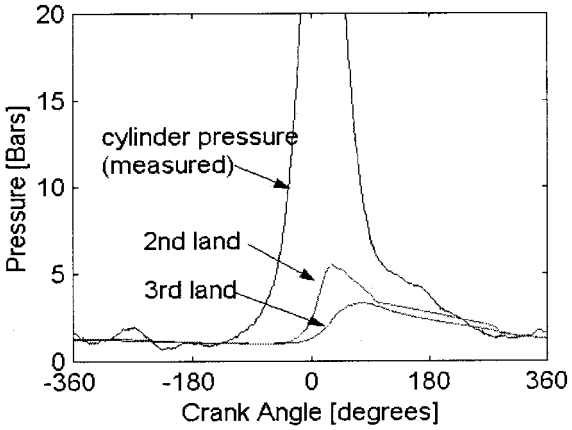


Fig. 8 Calculated land pressures under 6000 rpm/WOT

parable to each other, giving the possibility to real dynamics—flutter for the top ring around TDC of the compression/expansion strokes.

It was also found in [7,8] that the onset of top ring flutter was determined by static relative angle between the top ring and its groove. However, in this particular case, the onset of the top ring

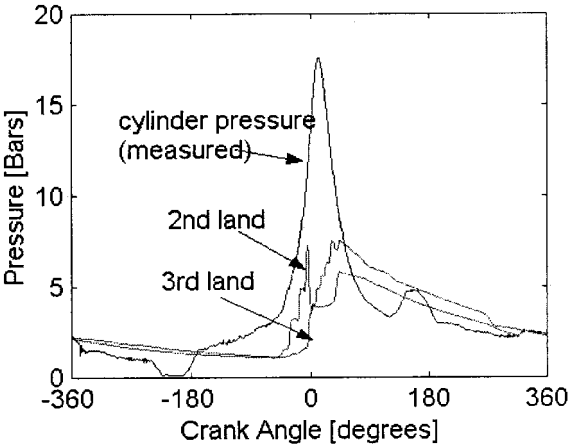


Fig. 9 Land pressures under 6000 rpm/no-load

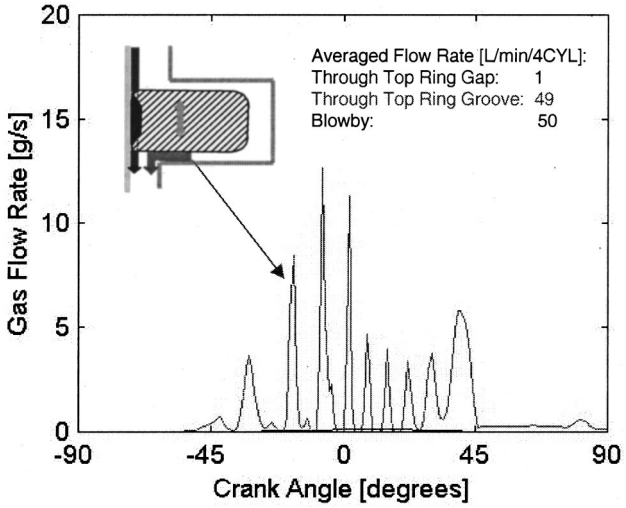


Fig. 10 Instantaneous and averaged gas flow rate through the top ring

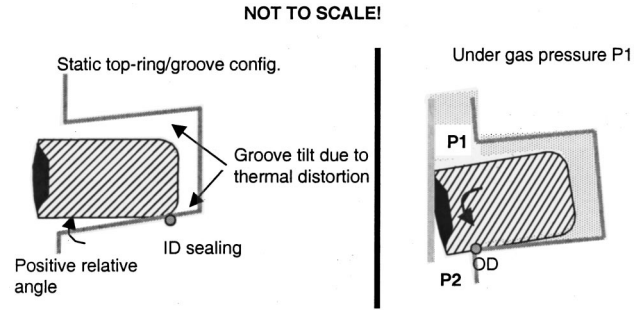


Fig. 11 Static top-ring/groove configuration and dynamic situation under 6000 rpm/no-load condition

flutter is less sensitive to ring/groove relative angle than the case studied in [7,8]. In this engine, the axial width of the top ring is 1.2 mm, which gives much less torsional stiffness than the 1.5 mm top ring used in [7,8]. Additionally, engine speed in this case is 6000 rpm, compared to 4000 rpm used in [7,8]. As a result, in this case, the rising cylinder pressure during late part of the compression stroke when the top ring still sits on the lower flank of the groove is able to twist over the top ring around the OD corner of the top ring groove (Fig. 11). Then, the gas with cylinder pressure penetrates into the lower flank of the top ring and creates ring lift. Consequently, top ring flutters despite the fact that a small static positive relative angle exists between the top ring. However, as will be shown later, a greater relative angle between the top ring and its groove can prevent top ring flutter from happening.

Under WOT, around TDC of compression/expansion strokes, Fig. 7 shows that the axial location of the CG of the top ring is elevated because the top ring is twisted over around the OD corner of the lower flank. However, contrast to the no load condition, the top ring is still kept in the vicinity of the lower flank of the groove. This is because high level of top land pressure or more precisely the gas pressure difference acting on the ring flank area outside of the OD corner of the top groove is able to overcome the upward inertia force of the top ring. Therefore, when the cylinder pressure is high, the only path for the gas to pass the top ring under WOT is the top ring gap.

Clearly, flutter of the top ring requires low cylinder pressure and high engine speed. Top ring flutter during no load condition and the resulted extra leakage through the top ring groove (Figs. 7 and 10) are the reasons of greater blow-by at no load than WOT at 6000 rpm seen in Fig. 5.

It is interesting to also examine how the second ring seals the large amount of gas flow through the top ring groove under the 6000 rpm/no-load condition. The second ring used in this engine is a standard Napier ring, and thus it has a positive static twist. Figure 12 shows the gas flow rate via different paths through the second ring. One can see sudden rise of gas flow rate through ring/liner interface due to the radial collapse of the second ring. As a result, gas directly flows through the interface of the second ring and the liner and causes sudden drop of the second land pressure and sudden rise of the third land pressure (Fig. 9).

Second ring collapse is an important phenomenon in high-speed engines. Although this paper focuses on the top ring, it is worthwhile to briefly describe the conditions for second ring radial collapse to occur. The principle of top ring collapse to be discussed later is essentially the same as that of the second ring. Second ring collapse can occur under the following conditions. First, the engine speed has to be sufficiently high so that the second ring is lifted at late part of the compression stroke due to high upward inertial force and slowly rising second land pressure. Second, the second ring has to have a positive static twist to form OD sealing with upper flank of the groove so that the second ring is stable after it is lifted. As a result of these two conditions, rising second land pressure acting on the taper face during late part of the com-

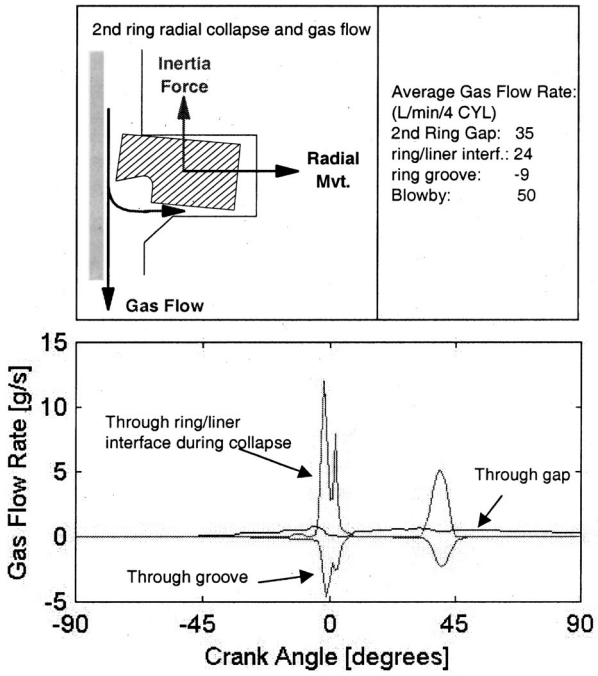


Fig. 12 Gas flow rate through the second ring under the 6000/ no-load condition

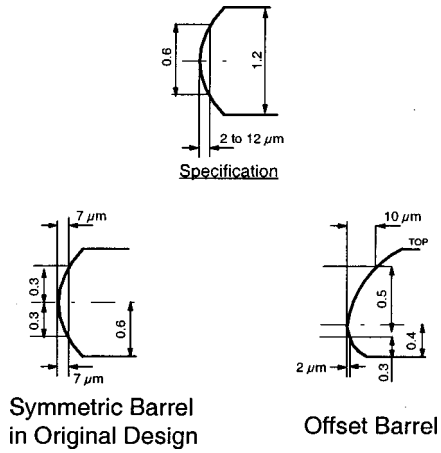


Fig. 13 Different top ring barrel faces

pression and early part of the expansion stroke can become great enough to overcome the ring tension and lower gas pressure in the second ring groove. Then, the second ring can be pushed inward, and radial collapse and direct leakage through the ring/liner interface can occur. Ring collapse gives another type of direct gas flow all over the circumference. Compared to ring flutter where the gas has to flow through the ring groove, ring collapse gives more direct gas flow with less resistance because the axial width of the ring is much shorter than that groove radial width.

Another important issue related to ring collapse is that ring collapse is more likely to occur when large area on the ring face is exposed to upper land gas pressure. Therefore, comparing different second ring designs, ring collapse should have less chance to occur with Napier rings than simple taper faced rings with a positive static twist.

It can be concluded that with the initial piston and ring designs, the gas sealing ability of the top two rings fails at the 6000 rpm/ no-load condition, due to top ring flutter and second ring collapse, resulting in extremely high blow-by.

Top Ring Radial Collapse and Flutter. In this SI engine, depending on the top ring and its piston groove design, top ring collapse could take place additional to flutter under high-speed/ light-load conditions. For instance, offset barrel face type of the top ring has been often used in this type of engine to avoid top ring's scraping up oil from the liner. Here, it can be demonstrated from the calculation that changing the initial symmetrical barrel face to offset barrel face (Fig. 13) would result in top ring collapse and even greater amount of blow-by under the 6000 rpm/no-load condition.

Figures 14 and 15 show the results of a calculation where only change compared to the initial design is an offset top ring barrel face was assumed (Fig. 13). One can see that at the 6000 rpm/no-load condition, additional to top ring flutter, radial collapse also occurs. Consequently, the blow-by value is more than doubled compared to the original design. The reason for the emergence of ring radial collapse is that the offset barrel face leaves greater face area exposed to the cylinder pressure than the symmetrical barrel face. Then the net inward force from gas pressures is increased, resulting in radial collapse.

Thus, in practice, one also has to pay attention to top ring's gas sealing ability when one intends to change the top ring face profile for altering ring/liner lubrication.

Blow-by Under 6000 rpm/WOT. As shown in Fig. 5, the engine blow-by at the WOT condition is also relatively high com-

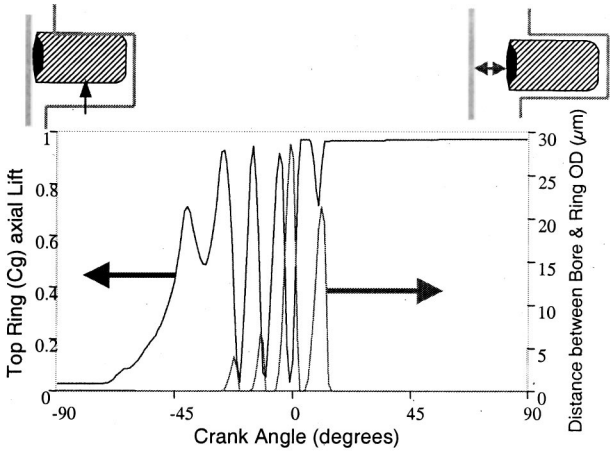


Fig. 14 Simultaneous top ring flutter and collapse with an offset barrel face under 6000 rpm/no-load

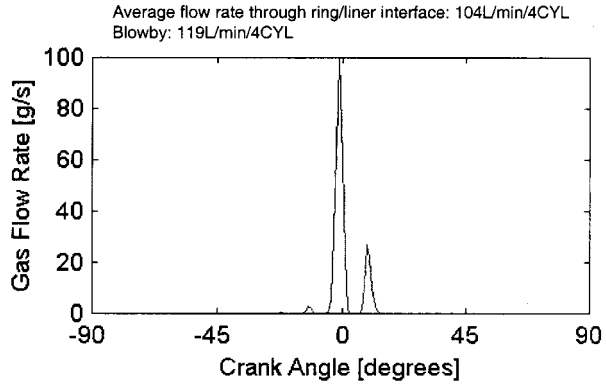


Fig. 15 Gas flow rate through top-ring/liner interface during top ring collapse

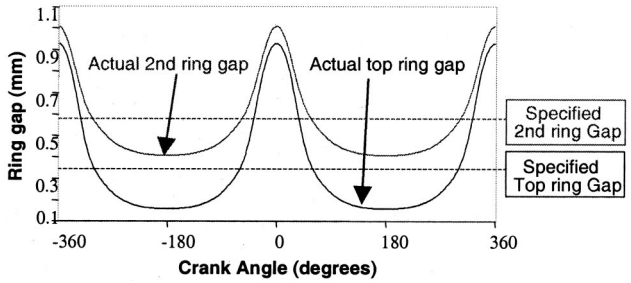


Fig. 16 Actual gap sizes of top two rings under WOT

pared to other production SI engines of this size. As demonstrated earlier, under WOT, top ring instability is not the cause of high blow-by. Rather, the reason is high bore expansion (Fig. 3), which is one of the characteristics of open deck aluminum blocks. With the bore expansion shown in Fig. 3, actual top ring gap as well as the second ring gap under WOT can become very large at TDC, as shown in Fig. 16. As a result, blow-by reaches fairly high value at 6000 rpm/WOT: 28 L/min. Figure 17 shows that almost all the blow-by is through the ring gaps. Additionally, in the initial design the second ring gap was large—a conventional way to control oil consumption.

Optimizing Piston and Ring Designs to Reduce Blow-by With all the understanding of main mechanisms that cause high blow-by and oil consumption, an attempt was made to optimize the blow-by (Fig. 5) and the oil consumption performance (Fig. 6) of this engine. Several modifications on piston and top two ring design (Fig. 18) were proposed and examined with assistance of the computer models before engine test. In the following, the proposed modifications are listed first. Then, results from calculation are shown before presenting the final engine test results.

To reduce the level of blow-by under high-speed/no-load engine running conditions, two types of top ring instabilities, namely, top ring flutter and radial collapse had to be minimized. Top ring modifications were the following.

(1) To avoid top ring flutter, sealing between the lower flanks of the top ring and groove was created (positive relative angle). Two options were possible. First, creating a negative tilt on the piston groove. However, when engine load increase, this change could generate top ring's scraping up the oil from the liner during late part of compression stroke and increase the oil consumption [9]. The second option is designing a top ring twist chamfer to create positive static twist. Additional to stabilizing the top ring, a top ring with a positive static twist could enhance top ring's down-scraping ability and reduce up-scraping, which is especially important for decreasing break-in oil consumption. Thus, the second option was chosen. As discussed earlier, the positive relative

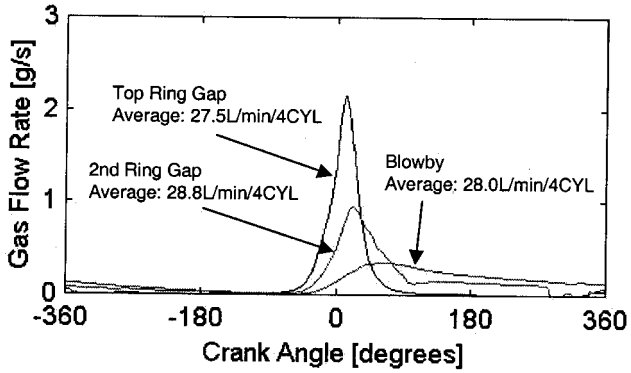


Fig. 17 Main gas flow and blow-by under 6000 rpm/WOT

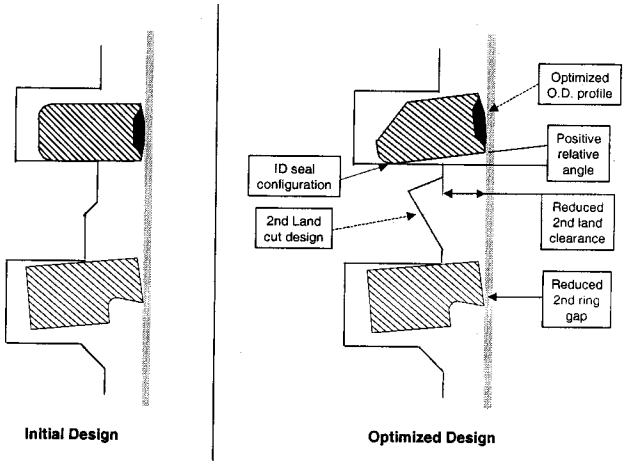


Fig. 18 Original and optimized designs

angle between the top ring and its groove has to be sufficiently large to be effective due to high engine speed and less top ring torsional stiffness.

(2) Symmetric barrel OD face profile was chosen to compromise between top ring collapse and up-scraping.

To reduce the level of blow-by under WOT the modifications were the following.

(1) Piston second land clearance was reduced to limit the gas flow through the top ring gap and to reduce second land pressure for avoiding second ring radial collapse.

(2) Second ring gap was reduced to limit the gas flow through the second ring.

(3) Special second land cut was created on the piston. This second land cut was proved to be critical to gas flow and oil transport in this engine, especially with reduced second ring gap. In terms of gas flow, the role of the second land cut is to avoid rapid second land pressure built up by increasing the inter-ring volume. As explained earlier, high second land pressure would create second ring collapse and increase gas flow through the second ring under high speed operating conditions. Moreover, second ring collapse can result in ring butting and break the ring. In terms of oil transport, the shape of this second land cut is intended to prevent oil from moving up along the second land by the inertia force and encourage oil transport from the second land to the liner.

Then, the computer models were used to optimize these parameters. Shown in the following are the calculation results based on the optimized piston and ring parameters.

Figure 19 shows the dynamics of the top ring with the optimized parameters under 6000 rpm/no-load. With the new design, the top ring is twisted over the groove's OD corner much later when the cylinder pressure becomes high during the compression stroke. Then, high cylinder pressure is able to hold the top ring in the vicinity of the lower flank of the groove in a similar manner as under the WOT. Therefore, these calculations show that the new design almost completely eliminated top ring flutter under the 6000 rpm/no-load condition. Additionally, keeping the top ring on the lower flank of its groove prevents radial collapse from occurring.

As a result of suppressing top ring flutter, the engine with the new design produces only 10 percent of the blow-by with the original design under the 6000 rpm/no-load condition. With slight flutter as seen in Fig. 19, the averaged gas flow rate through the top ring groove is reduced to 3 L/min. Certainly, greater static twist would narrow the conditions for the top ring flutter to occur and further suppress top ring flutter. However, the top ring static twist is limited by the fact that it might block the groove to induce ring collapse.

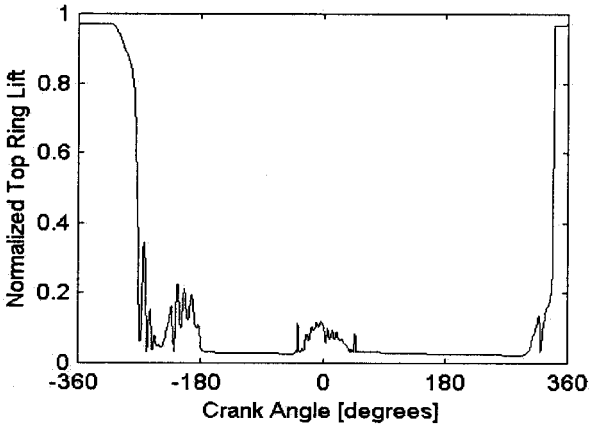


Fig. 19 Top ring dynamics under 6000 rpm/no-load with optimized design

Under 6000 rpm/WOT (Fig. 20), the main top ring behavior change occurs around TDC compression/expansion where the top ring is twisted over around the OD corner much later and less severely with the new design because of the positive static twist with the new design. Although this subtle change of ring dynamics has no impact on blow-by, it certainly reduces the severity of the contact at the OD corner between the lower flank of the top ring and its groove.

The calculations show that blow-by under 6000 rpm/WOT was reduced with the new design. Due to the decrease of the piston second land clearance, the averaged gas flow value through the top ring gap was reduced from 27.5 to 18.8 L/min. Because of the smaller second ring gap and the increased second land volume, the averaged gas flow rate through second ring gap was reduced from 28.8 to 17 L/min. These two improvements result in a 32 percent engine blow-by reduction (28–19 L/min).

Figure 21 plots the measured blow-by map (Fig. 21a) of the engine with the optimized configurations. Compared to the initial blow-by performance shown in Fig. 5, one can clearly see the remarkable improvements over all the speed and load conditions with the new designs. In Fig. 21b, one can also see the impact of the blow-by improvement on the ''transient'' oil consumption, measured in a fleet test with ''sport'' driving. The engine transient oil consumption with the new designs is only 10 percent of the original one.

This example illustrates the advantages of using computer models to understand in details the physics involved in modern SI engines. In this case, by no means, the modifications were conventional. However, with the help of model analysis, number of engine test iterations and thus development time were substantially shortened.

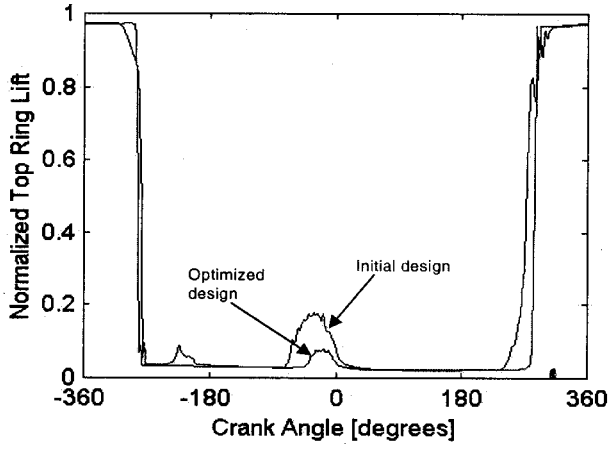
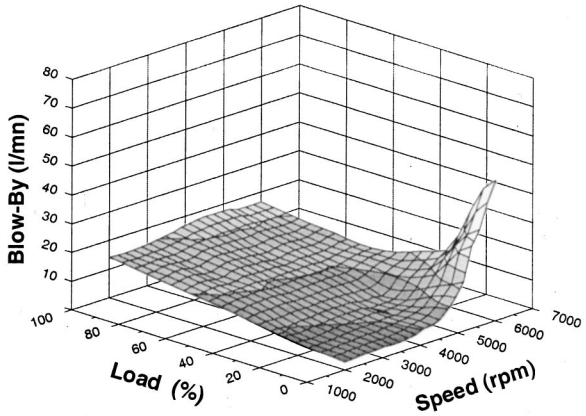
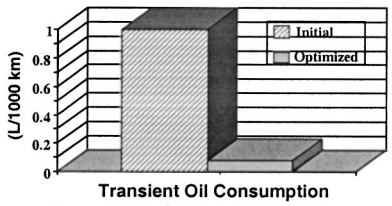


Fig. 20 Comparison of top ring dynamics under 6000 rpm/WOT of two designs



21a. Blowby Map with the Optimized Design.



21b. Comparison of Transient Oil Consumption.

Fig. 21 Results with the optimized design

Reverse Flow Through Top Ring Flutter

It has been widely accepted that reverse gas flow to the combustion chamber from the second land is one of the main mechanisms for oil consumption. The reverse flow through the top ring may take place via two paths, namely, through the top ring groove and top ring end gap. It was argued by the authors in a study of a passenger car diesel engine [9] that main contributor to the oil consumption through reverse gas flow is the reverse flow through the top ring groove when the top ring flutters during late part of the expansion and early part of the exhaust. The same conclusion was drawn in some of our HD diesel engine studies. Here, an example is given in a SI engine where reverse flow through the top ring groove is also considered to be the dominant path for the oil consumption through gas stream.

This example is based on a 2.0 L SI engine. Again, the same models are used to conduct the analysis. Figure 22 shows the dynamics of the top ring and Fig. 23 shows the land pressures. One can see that the top ring flutters during early part of the exhaust stroke, which induces sudden drop of the second land pressure. This top ring flutter is called ''reverse flutter'' by the authors because all the forces and gas flows are completely in the direction opposite to the one of the other type of top ring flutter described earlier in this paper. During early part of the exhaust stroke, first of all, inertia force is downward. Second, the second land pressure is higher than the cylinder pressure. And third, the lower flanks of the top ring and the groove form an ID sealing because of the combination of the top ring twist and downward tilt of the groove due to thermal deformation.

When this reverse flutter occurs, the corresponding reverse flow through the groove is the primary mechanism for blowing the oil on the piston into the combustion chamber due to the following two reasons. First, the reverse gas flow rates through the top ring

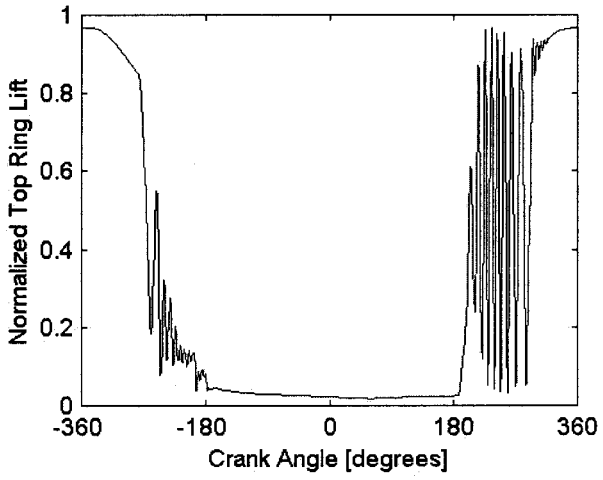


Fig. 22 Top ring dynamics under 4000 rpm/WOT

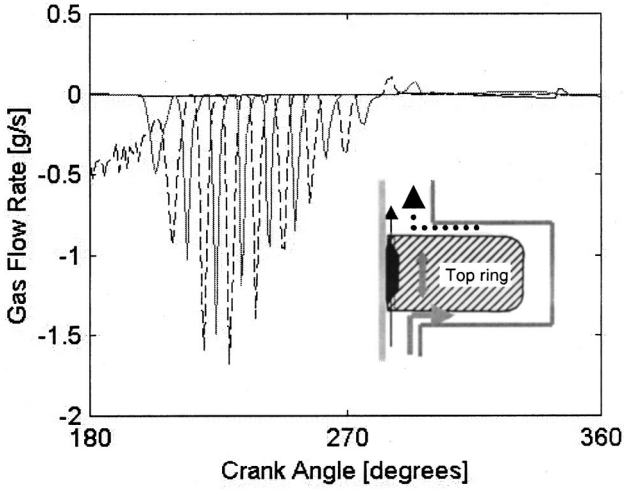


Fig. 24 Gas flow rate via different routes through the top ring during the exhaust stroke

groove are much higher than the one through the top ring gap (the gas flow rate through the top ring gap is so small that it is invisible in Fig. 24), as can be seen from Fig. 24. Second, as the reverse flutter occurs all over the circumference, there is more available oil for the reverse gas flow to carry. One has to realize that with appreciable engine load, the main flow through the top ring gap is downwards and thus the second land region below the top ring gap is cleaned [10]. Therefore, there is no oil for the reverse flow through the top ring gap to carry. In contrast, oil on the second land and inside the top ring groove all over the circumference is available for the reverse flow through the top ring groove to entrain [10,11].

Despite a clear understanding of reverse flutter phenomenon, it is difficult to avoid it. ID sealing between the lower flanks of the top ring and its groove usually exists because of groove thermal distortion, wear, and positive top ring static twist that is required often for other purposes as described earlier. Thus, one can see the complexity of ring designs. One parameter could be beneficial to one process and detrimental to another one. One has to react differently with different engines and engine operating characteristics.

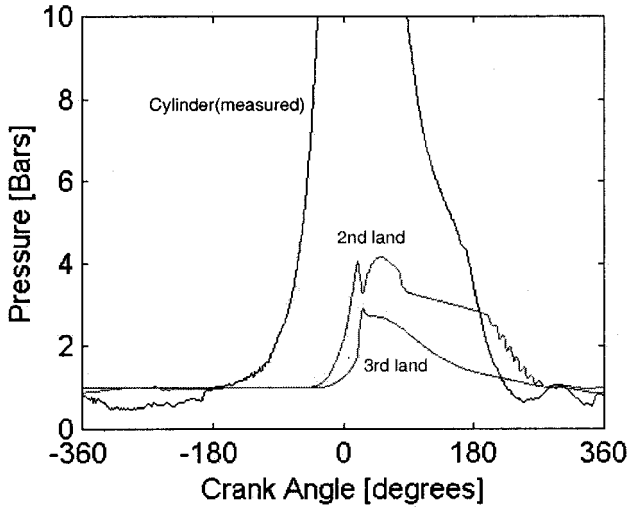


Fig. 23 Land pressures under 4000 rpm/WOT

Top Ring Severe Lubrication

One of the most critical factors affecting SI engine development in recent years has been the need to improve fuel efficiency. Typically, the friction in the engine accounts for 8 percent of passenger car engine power [12]. One major effort for reducing engine friction is to formulate low-friction engine oils and it has been proved to be an effective way [5]. However, low-friction low-viscosity oils can also significantly modify and worsen the lubrication conditions between the top ring and the cylinder bore as well as between the top ring and piston groove. Thus an integrated effort has to be made from all aspects of engine lubrication system. With the trend of lowering oil viscosity and increasing component temperature, it is not surprising that in the past few years, top ring lubrication problems, namely scuffing and microwelding, became more and more frequent in SI engine development. Computer models enable us to conduct detailed analyses on ring/liner lubrication, and the contact between the rings and the liner as well as between the rings and the grooves. With these analyses, one can often discover the specific mechanical behavior that causes the failure and improve the design with simple geometrical changes instead of blindly testing new materials and coatings.

All the calculations in this section are conducted with the latest models developed in the MIT lubrication consortium. One of the major improvements in the ring pack lubrication model since the last publication [13,8] is that the model considers the effects of ring dynamic twist and piston dynamics tilt on ring/liner lubrication.

Top Ring/Bore Lubrication. In modern SI engines, top ring/bore lubrication occurs in a hostile environment characterized by high temperatures, high gas pressures, acidic combustion products, and thin and highly loaded oil film. The top ring experiences all three different kinds of lubrication modes:

(i) Hydrodynamic lubrication where the oil viscosity is the determining factor for friction power losses.
(ii) Boundary lubrication where the surface material, roughness structure, and oil friction modifiers are the determining factors for friction and wear.
(iii) Mixed lubrication.

Considering the lubrication between the top ring and the liner, the top ring may experience boundary lubrication and severe asperity contact when it travels in top of the liner. At high load engine running conditions, boundary lubrication and severe asperity contact occur on the top ring around TDC of the compression/

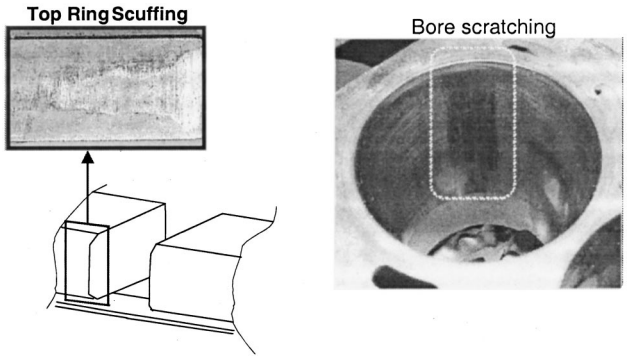
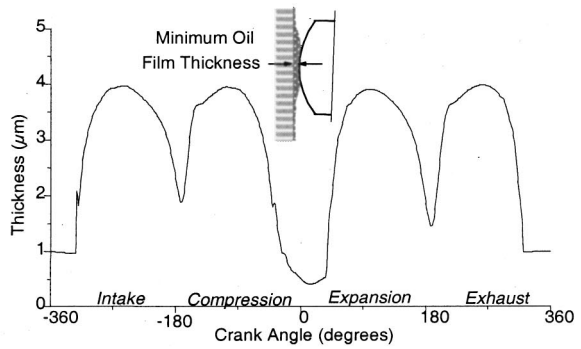
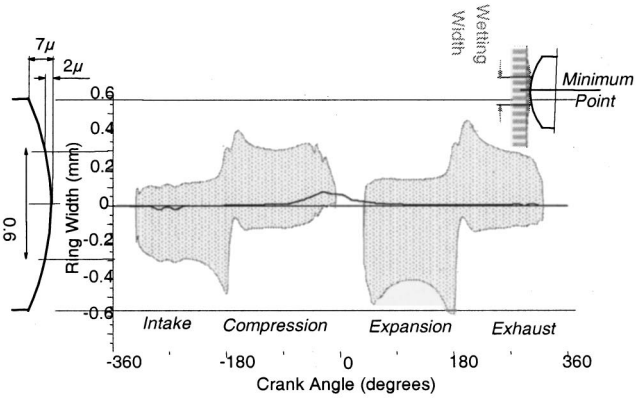


Fig. 25 Top ring and bore scuffing



(a) Minimum Oil Film Thickness of the Top Ring



(b) Top Ring Wetting Condition

Fig. 26 Lubrication between top ring and liner

expansion strokes because the high gas pressure pushed the ring outward from the top ring groove. Around TDC of the exhaust/intake strokes, the top ring usually experiences pure boundary lubrication because the amount of oil left on the top of the liner after expansion stroke is extremely small [13]. Yet the magnitude of the resulting asperity contact is not significant due to lack of existence of high gas pressure and thus not important for generating scuffing. Therefore, in a warm engine, top ring scuffing (Fig. 25) is most likely initiated at TDC of the compression/expansion strokes when the bore and the top ring coating and/or material can not survive high level of asperity contact.

Although it appears to be a material failure when scuffing occurs, the root of scuffing may be bad mechanical behavior of the rings with their mating parts and a simple change of piston and ring geometry may effectively improve the mechanical behavior and avoid scuffing from occurring. To illustrate the effectiveness of geometrical changes, computer model [9] was used to study top ring/bore lubrication in a 1.8 L SI engine with a cast iron block under 5000 rpm/WOT. With 5W20 oil, this engine initially encountered some scuffing problems (Fig. 25) that were then solved by improving ring, liner, and cooling designs. For illustration purposes, the influences of three major parameters on the asperity contact between the top ring and the liner were investigated: oil SAE viscosity grade, top ring barrel face profile, and bore surface roughness.

All the parameters are based on design parameters and major parameters for the baseline condition are listed as follows:

(i) Top ring width: 1.2 mm.
(ii) Roughness: 0.1 μm in Rq for the barrel face and 0.5 μm in Rq for the liner.
(iii) Oil: SAE 5W20 with HTHS viscosity of 3.2 mPa at 150°C.
(iv) Liner temperature: 100°C at TDC and 80°C at the BDC.
(v) Top ring running surface profile: 2 μm of barrel drop—lower limit of the specification, as shown in Fig. 26.

Baseline results are plotted in Figs. 26 and 27. In Fig. 26, the minimum oil film thickness (MOFT) of the top ring (a) and wetting condition (b) are plotted. The results in Fig. 26 are for the ring section along the anti-thrust side because the piston is designed with a pin offset to the thrust side and dynamic piston tilt makes the top ring experience most severe lubrication condition at the anti-thrust. For this matter, all the rest results are also for the ring section along the anti-thrust side. In Fig. 26b, shaded area represents the region on the top ring running surface where oil is present between the ring and liner and hydrodynamic pressure is generated [13,8]. Thus, pure boundary lubrication occurs if there is no shaded area in Fig. 26b, e.g., around the TDC. As explained in [13,8], oil supply to the liner region above the TDC position of the oil control ring does not necessarily repeat from cycle to cycle. Here, the results show that there is no hydrodynamic lubrication in the region above the TDC position of the oil control ring. However, hydrodynamic lubrication may return in another cycle because the top ring is bale to carry some oil to that region. Whichever cycle is chosen won't affect the conclusions to be drawn later.

Figure 27 shows the asperity contact pressure distribution between the top ring and the liner around TDC of the compression/expansion strokes. The maximum contact pressure value is about 200 bars and boundary lubrication takes place for about 60 deg of crank angle. This type of top ring lubrication condition shown in Figs. 26 and 27 is representative of the engines in the market using low-viscosity lubricants and can partially illustrate the requirements of the top ring coating evolution.

Effects of Lubricant Types on Friction and the Contact Between the Top Ring and the Line. A thicker oil, SAE 10W50,

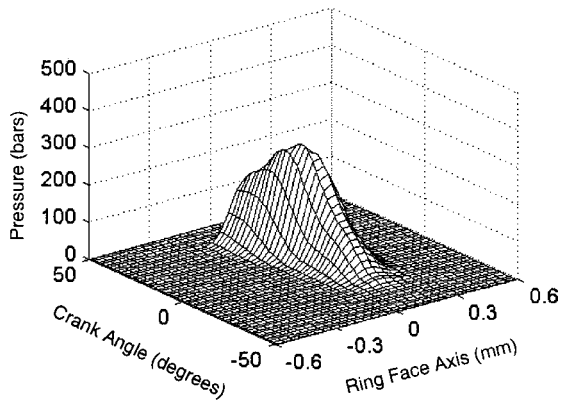


Fig. 27 Top-ring/bore asperity contact pressure

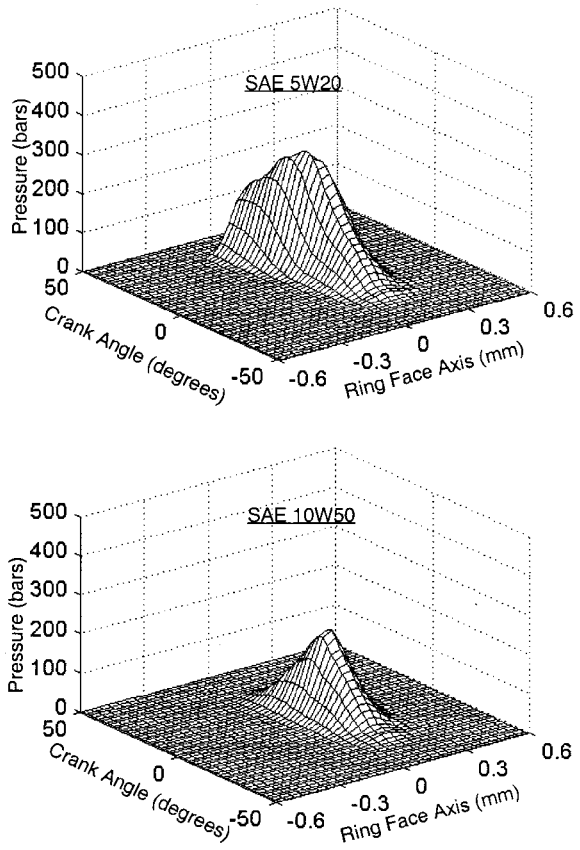


Fig. 28 Influence of oil properties

was chosen to illustrate the impact of the oil viscosity evolution on top ring lubrication. The HTHS viscosity of this SAE 10W50 at 150°C is 3.42 mPa s. Other parameters were kept the same as the baseline condition.

Figure 28 shows the comparison of the contact pressure between SAE 10W50 and SAE 5W20. One can see that the magnitude of asperity contact pressure is decreased by 25 percent. Additionally, with SAE 10W50, asperity contact only lasts about 20 deg of crank angle.

Table 1 shows the calculated friction mean effective pressure (FMEP) with these two different lubricants. One can see that the FMEP of the entire ring pack with SAE 5W20 is approximately the same as the one with 10W50. Further examining the FMEP contribution from different rings and different mechanisms in Table 1, it can be seen that with 5W20, higher friction contribution from boundary lubrication on the top ring offsets all the friction benefits obtained from the second ring and the oil control ring.

It has to be stated that in this study, these two oils were assumed to have the same friction modifiers and to have 0.1 as the

Table 1 Influence of oil viscosity on ring pack friction power losses

	5W20 configuration	10W50 configuration
Top ring 'boundary" FMEP (kPa)	3.0	0.9
Top ring FMEP (kPa)	6.3	4.9
Ring Pack FMEP (kPa)	17.5	17.6

Table 2 Influence of oil viscosity on ring pack friction power losses at a higher wall temperature (20°C)

	5W20	10W50
Top ring 'boundary' FMEP (kPa)	3.0	3.0
Top ring FMEP (kPa)	5.5	5.7
Ring Pack FMEP (kPa)	14.0	15.0

friction coefficient of the boundary lubrication. Application of friction reduction additives in 5W20 could reduce friction coefficient to 0.05 [5] and would indeed decrease friction power losses with 5W20. In addition, the liner temperatures used may be too low compared to the engines with aluminum blocks. To show how the performance of these two lubricants compare at higher liner temperature, another calculation was made by assuming that the liner temperature was 20°C higher than the baseline. One can see the advantage of SAE 5W20 on friction from Table 2. Clearly, at higher temperature, using SAE 10W50 results in the same kind of oil starvation around the TDC as 5W20.

This study shows that with low-viscosity oils, the top ring spends longer time experiencing boundary lubrication than with higher-viscosity oils. Consequently, under certain conditions, low-viscosity oils may actually produce more FMEF in the piston ring pack than higher viscosity oils. Thus, low-viscosity oils must be blended with better friction modifier in order to effectively reduce friction. From the point of view of piston ring designs, there is possibility that the magnitude of asperity contact pressure can be reduced with better ring face profile and surface roughness.

Effects of the Top Ring Running Surface Profile and Surface Roughness. As mentioned earlier, in the baseline case, lower limit of the barrel face drop thus a flat surface profile was used. It is interesting to examine how ring/liner contact varies within the tolerance of the barrel face design by using the average barrel drop 7 μm, as specified in the design. With a small axial width at 1.2 mm, this change of barrel face greatly increases the curvature of the face profile and sharpens the top ring face. Figure 29 shows the comparison of top-ring/liner contact pressure distribution. One can see that sharper face profile with 7 μm barrel drop results in higher magnitude in contact pressure because of less contact area.

Asperity contact of a sharp top ring can bring very localized and hard contact between the top ring and the bore. Then, the heat generated by localized high friction may not be fully released during the cycle, resulting in local temperature increase and thermal expansion. In turn, thermal expansion produces sharper surface (thermal bump) and thus results in more localized heat generation and temperature increase. As a result, under certain circumstances, local temperature becomes out of control. When the local temperature becomes sufficiently high, local microwelding occurs on the top ring (Fig. 25). Microwelding increases the hardness of the running surface and sharpens the running surface, and further development would produce severe scoring or scratching of the bore (Fig. 25). This phenomenon is classified as thermalelastic instability and is also the cause of failure for many other lubrication systems, e.g., mechanical face seals [14].

Although wear is generally not considered to be a favorable feature for materials, proper wear can remove the thermal bump and thus prevent the thermalelastic instability from happening on the top ring face. On the other hand, a top ring with high wear resistance coating/treatment may enhance thermal elastic instability because of its inability to remove the thermal bumps.

The result of top ring scuffing with hard coatings is a rapid increase of the engine oil consumption. On the other hand, a top ring coating with a lower wear resistance may result in more wear

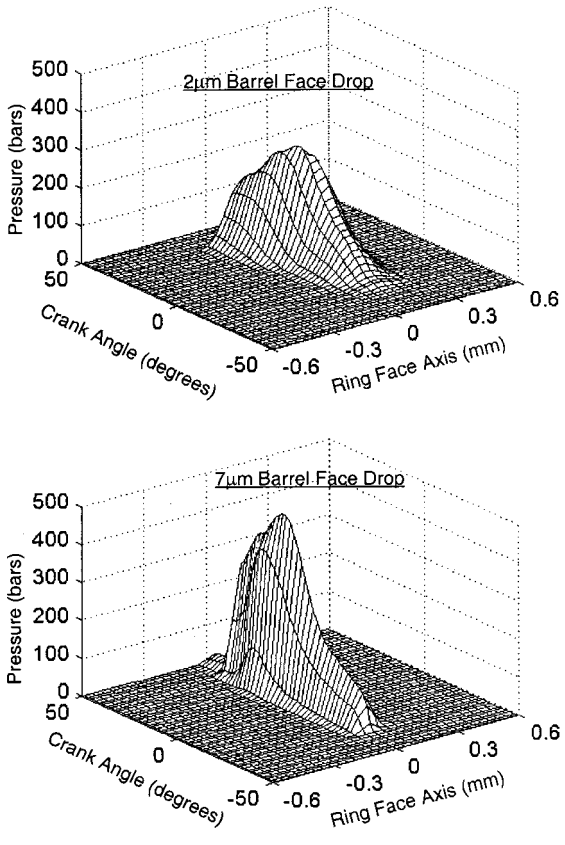


Fig. 29 Influence of barrel profile

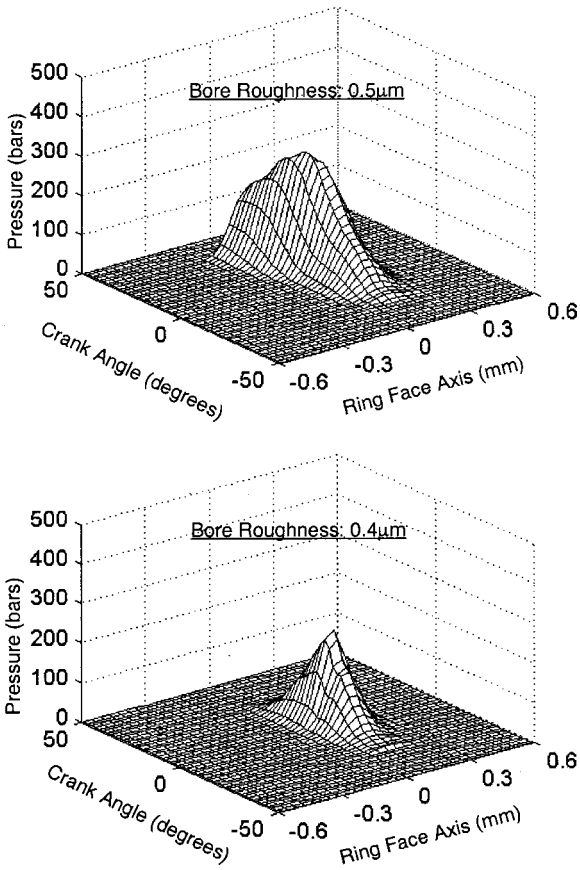


Fig. 30 Influence of liner roughness

on the top ring running surface, and thus top ring gap and blow-by. Additionally, significant wear on the top ring OD face flattens the surface profile and thus may increase top ring scraping and oil consumption. Thus, in reality, one has to find a compromise between wear resistance and scuffing resistance.

Other important features to consider for reducing friction and contact between the top ring and liner include self-lubrication properties of the coatings [15] and surface roughness/finish. Figure 30 shows the comparison between the baseline and an assumed smoother liner. One can see a dramatic decrease in the magnitude of contact pressure and contact area because hydrodynamic lubrication is enhanced with lower surface roughness.

As a final note, the initial scuffing problem that occurred in this 1.8 L cast iron block engine was solved by reducing the tolerance of both barrel face profile and bore surface finish, and by improving the efficiency of the engine cooling system.

Top Ring Piston Groove Wear and Microwelding. In modern SI engines, light weight piston design, shorter top land height, and increase of engine performance all lead to high temperature on the surface of the top ring groove (up to 260°C). Under high groove surface temperature, high contact pressure between the top ring and its groove can generate local fusion between the top ring and the groove. This effect, commonly called microwelding, is in fact a metal transfer from the groove to the ring. Usually, as a result of microwelding, piston groove is destroyed, and engine blow-by and oil consumption are greatly increased. Figure 31 is an example of microwelding in a SI engine where the lower flank of the groove was damaged and aluminum was welded into the lower side of the top ring.

Although piston material and coating [3] as well as treatment on the top ring flanks [15] could be effective ways to avoid microwelding, they could be costly. On the other hand, some geometric design changes of the top ring and its piston groove can effectively reduce the contact pressure between them, and sometimes solve the microwelding problem. Certainly, sometimes optimizing geometry and minimizing contact pressure may still not be able to avoid microwelding. Then, one can consider other solutions in coatings, piston material, and surface treatment. Another solution for microwelding is bringing more oil into the top ring groove to cool the groove and to enhance hydrodynamic pressure between the flanks the top ring and groove. However, by doing so, more oil consumption and degradation would be generated.

To illustrate how the magnitude of the contact pressure can be reduced between the flanks of the top ring and the groove, a simple example is given using two different piston tilt values. Again, computer models [7,8,9] were used. Figure 32 shows the asperity contact pressure distribution between the lower flank of the top ring and its groove. Figure 32 only shows the results around the TDC of the compression stroke where the top ring is twisted over around the OD corner. One can clearly see the re-

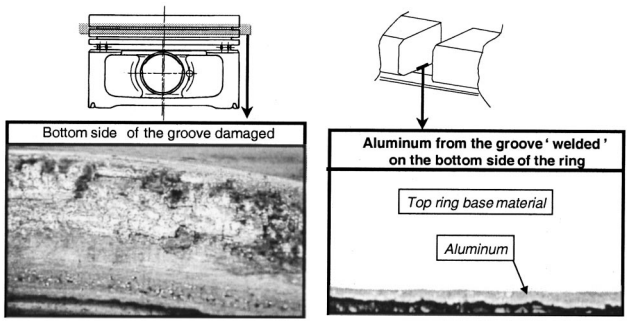


Fig. 31 Top-ring/piston-groove microwelding

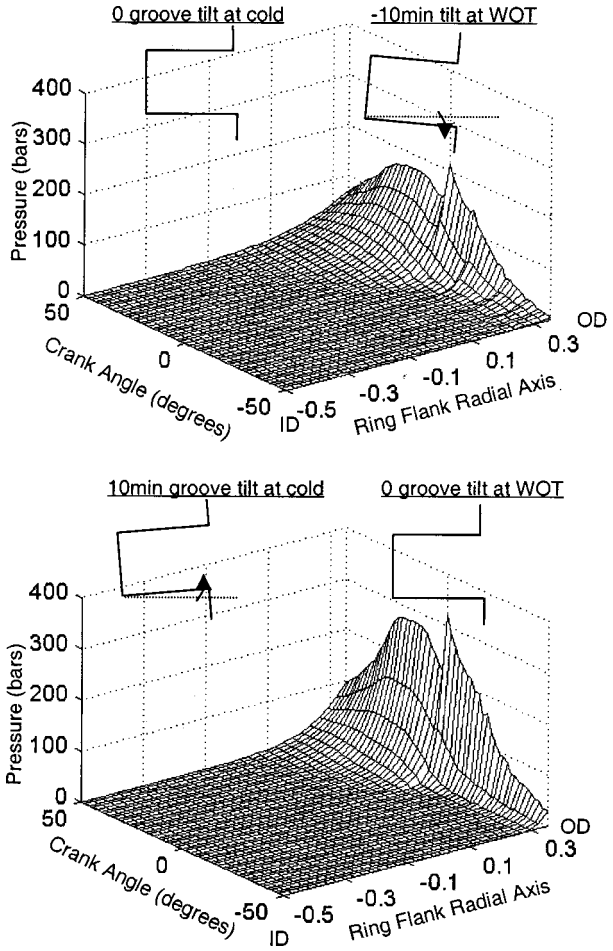


Fig. 32 Influence of the piston groove tilt on the contact pressure distribution between the lower flank of the top ring and the groove

duction of the magnitude of the asperity contact pressure with the groove that does not have an up-tilt. As discovered in an earlier work [9], the wear of top ring and the groove tends to produce a relative angle between the lower flanks of the top ring and the groove such that during engine operation, the asperity contact pressure distributes uniformly in the radial direction between the lower flank of the top ring and the groove. Certainly, if the groove tilt angle is further reduced to a negative value, the peak contact pressure at the OD corner can be further reduced. As one knows, a negative top groove tilt may result in excessive top ring up-scraping. Therefore, based on the considerations of top-ring/groove contact and top ring scraping, an up-tilt top ring groove combined with a large top ring static twist may produce less asperity contact pressure between the lower flank of the top ring and the groove without increasing top ring up-scraping. By doing so, wear and microwelding between top ring and the groove as well as oil consumption may all be reduced.

Conclusions

Through practical examples, this paper shows several critical aspects of top ring design in modern SI engines. It was found that many engine design and engine operating characteristics in modern SI engines bring new challenges to top ring design in controlling engine blow-by, friction, wear, and oil consumption. High engine speed creates more ring instability in both axial and radial directions, and flutter and ring radial collapse generate more direct and stronger gas flow all around the circumference of the piston than otherwise only through ring gaps under low speed conditions. Thus, the main routs for blow-by and oil consumption become different from low speed conditions, and design strategies to control blow-by and oil consumption under high speed conditions have to be adjusted. Additionally, at high engine load conditions, high bore expansion in engines with aluminum blocks gives significant increase to ring gaps when rings travels at top of the liner. To control blow-by, one has to make adjustment on the second land and the second ring in order not to give a direct path to the increased gas flow through the top ring gap. The computer models used are helpful in identifying the main mechanisms contributing to increased gas leakage and gas flow related oil transport under high speed conditions and in optimizing piston and ring parameters to minimize blow-by.

Increase of piston temperature and adaptation of low-friction oils in modern SI engines bring more and more problems in micro-welding between the top ring and the groove as well as scuffing between the top ring and the liner. While changes in materials, coatings, and surface treatment may ultimately be needed, magnitude of asperity contact pressure between the top ring and its mating parts can be reduced by making simple changes in the geometry of the piston and top ring. By reducing contact pressure between the top ring and its mating parts, the onset of microwelding and scuffing can certainly be delayed if not completely avoided. Thus, when microwelding or scuffing occurs, it is important to examine the mechanical behavior of the piston and rings before investing on material changes.

Acknowledgments

Tian Tian's work is sponsored by the MIT Lubrication Consortium in IC engines whose current members include Dana, Renault, Peugeot PSA, Volvo, and Mahle. The computer models were developed inside this MIT consortium. The authors would like to thank Alfred Schreiber of Dana Perfect Circle Europe (PCE) division for his expertise in the ring design and application areas and the technicians in the test benches and material lab at PCE for their quality work.

References

[1] Nakada, M., 1993, ''Piston and Piston Ring Tribology and Fuel Economy,'' Proceedings of International Tribology Conference, Yokohama.
[2] Hill, S. H., and Sytsma, S. J., 1991, ''A Systems Approach to Oil Consumption,'' SAE paper 910743.
[3] Dr.-Ing. Hartmut Kamp et al., ''Piston Technology for Passenger-Car Gasoline Engines, Lightweight Design—Microwelding—Direct Fuel Injection,'' Kolbenschmidt—Competence in Engines.
[4] Dr.-Ing. Eduard Koeler, 1995, ''Aluminum Engine Blocks,'' Kolbenschmidt special print for IAA.
[5] Hoshino, K., et al., 1998, ''Fuel Efficiency of SAE 5W-20 Friction Modified Gasoline Engine Oil,'' SAE paper 982506.
[6] Cheng et al., 1993, ''An Overview of Hydrocarbon Emissions Mechanisms in Spark-Ignition Engines,'' SAE paper 932708.
[7] Tian, T., Noordzij, L. B., Wong, V. W., and Heywood, J. B., 1998, ''Modeling Piston-Ring Dynamics, Blow-by, and Ring-Twist Effects,'' ASME J. Eng. Gas Turbines Power, **120**, pp. 843–854.
[8] Tian, T., 1997, ''Modeling the Performance of Piston Ring-Pack in IC Engines,'' Ph.D. thesis, MIT, Cambridge, MA.
[9] Tian, T., Rabute, R., Wong, V. W., and Heywood, J. B., 1997, ''Effects of Piston-Ring Dynamics on Ring/Groove Wear and Oil Consumption in a Diesel Engine,'' SAE paper 970835.
[10] Thirouard, B., Tian T., Hart, D. P., 1998, ''Investigation of Oil Transport Mechanisms in the Piston Ring Pack of a Single Cylinder Diesel Engine, Using Two Dimensional Laser Induced Fluorescence,'' SAE paper 982658.
[11] Min, B., et al., 1998, ''Dynamic Characteristics of Oil Consumption—Relationship between the Instantaneous Oil Consumption and the Location of Piston Ring Gap,'' SAE paper 982442.
[12] Shannon, I., and Wakem, M. P., 1999, ''Lubricants Matter,'' Engine Technology International issue 1/99.
[13] Tian, T., Wong, V. W., and Heywood, J. B., 1996, ''A Piston Ring-Pack Film Thickness and Friction Model for Multigrade Oils and Rough Surfaces,'' SAE paper 962032.
[14] Lebeck, A. O., 1991, ''Principles and Design of Mechanical Face Seals,'' John Wiley and Sons, New York.
[15] Shuster M., et al., 1996, ''Piston Ring Microwelding Phenomenon and Methods of Prevention,'' SAE paper 960745.

Inlet Losses in Counterflow Wet-Cooling Towers

E. de Villiers

D. G. Kröger
e-mail: dgk@maties.sun.ac.za

University of Stellenbosch,
Department of Mechanical Engineering,
Stellenbosch, 7600 South Africa

The inlet loss coefficients for dry, isotropically packed, circular and rectangular counterflow cooling towers are determined experimentally and empirical correlations are formulated to fit this data. Computational fluid dynamics is used to investigate the dependence of the inlet loss coefficient on the rain zone characteristics. The rain zone generally dampens the inlet loss, but the coupling is indirect and involves a large number of dependent variables. The numerical model is validated by means of experimental data for dry towers and it is found that the degree of accuracy achieved for circular towers exceeds that for rectangular towers. Consequently, the correlation derived to predict this occurrence for circular towers can be applied more confidently than its rectangular counterpart. [DOI: 10.1115/1.1359236]

Introduction

A method for predicting the thermal performance of cooling towers must include accurate representations of the tower's airflow resistances. Since inlet losses constitute a significant portion of these resistances, the importance of an accurate model to predict their magnitude is clear. Although some relations exist to describe these phenomena, they tend to be incorrect or limited to specific tower configurations. Furthermore, finite element methods that have the potential to accurately describe inlet losses [1] are not often used in practice since they tend to be relatively time consuming and are unsuited to optimization practices.

Many researchers [2–5] have investigated inlet losses in an attempt to quantify their influence on tower performance. The problem of applying their analyses to counterflow wet-cooling towers is twofold: (1) Most of the earlier researchers fail to simulate the fill (Fig. 1) in a manner consistent with operational tower packing. Terblanche and Kröger [6] show that inlet losses are highly dependent on fill configuration and characteristics. It is therefore important that the fill being used in an experimental investigation of tower inlet losses be geometrically similar to actual fills. (2) None of the previous analyses incorporate the effect of the rain zone resistance on the inlet loss. Since it has been shown that inlet losses are highly dependent on the magnitude of the fill resistance [6,7] it is safe to assume that the rain zone (which often has a considerable resistance) will also influence the size of the inlet loss. (This will later be shown to be the case.)

The purpose of this paper is, thus, to derive a method to predict the inlet losses for natural and induced draft wet-cooling towers that incorporates variable fill geometries and the rain zone resistance. The loss coefficient produced by this method is intended for use in a one-dimensional performance evaluation of the appropriate cooling tower.

Both experimental and numerical (computational fluid dynamics) approaches have been found inadequate when used to study inlet losses. Experimental methods, because of the impracticality of simulating the rain zone other than on a scale of one to one. CFD, because of current commercial code's inability to accurately simulate large scale eddies present within the recirculating region in cooling tower inlet flows. A solution based on a compromise between the two methods was therefore sought. Experimental procedures are employed to find the absolute value of the inlet loss while ignoring the effect of the rain zone. CFD is used to find the relative influence of the rain zone on the inlet loss which can then be applied to the experimental results to find an approximation of the true inlet loss.

Contributed by the Power Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the C&E Division, Jan. 28, 2000; final revision received by the ASME Headquarters Jan. 23, 2001. Editor: D. Lou.

Experimental Evaluation of Inlet Losses

The inlet loss coefficient for a counterflow cooling tower can be expressed in terms of the change in total pressure between stagnant ambient air far from the tower and the air after the fill assembly. Assuming a uniform velocity distribution after the fill assembly the inlet loss coefficient can be defined as

$$K_{ct}=\frac{p_a-(p_o+\rho_o v_o^2/2)-\int_0^{H_o}\rho g\,dz}{\rho_a v_i^2/2}-K_t\left(\frac{\rho_a}{\rho_{fi}}\right)\left(\frac{A_i}{A_{fr}}\right)^2 \quad (1)$$

where A_i is the inlet cross-sectional area of the tower, A_{fr} is the frontal area of the fill, and ρ_{fi} is the harmonic mean density of the air flowing through the fill. The subscript ''*i*'' refers to conditions at the cross-sectional inlet, ''*o*'' to conditions above the fill assembly, and ''*a*'' represents ambient conditions. Furthermore, z is the direction perpendicular to the surface, so that the integral represents the pressure drop caused by the change in elevation of the air flow. For a wet-cooling tower the resistance term, K_t, is the sum of the known loss coefficients between the tower entrance and the top of the fill assembly, all referred to fill conditions.

To obtain experimental values for the inlet loss coefficient in accordance with Eq. (1), certain simplifications and approximations have to be made. In addition to the fact that rain zone effects are ignored for the time being, the following assumptions were made concerning the experimental arrangement:

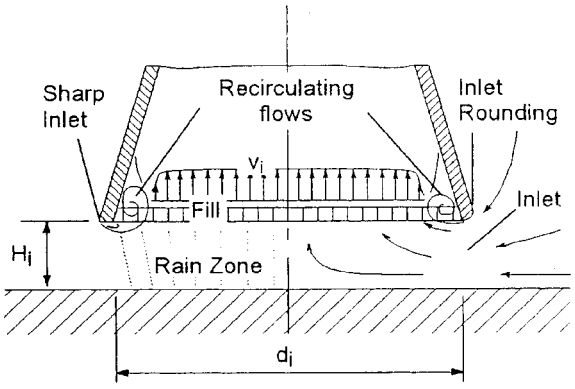


Fig. 1 Schematic representation of a cooling tower inlet

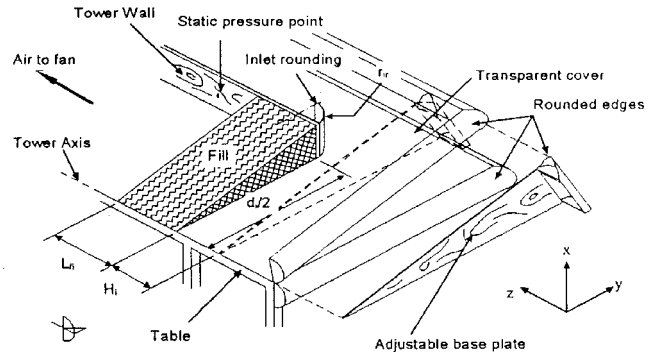


Fig. 2 Section model of a cylindrical cooling tower

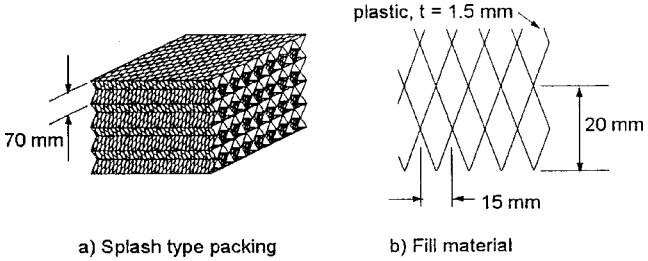


Fig. 3 Isotropic resistance fill material (trickle grid)

• The effects of gravity can be ignored. By using a horizontal test section, the integral term that represents the buoyancy force in Eq. (1) becomes redundant, thus $\int_0^{H_o} \rho g dz = 0$.

• The flow is adiabatic and incompressible. Since there is no heat or mass transfer in the scale model and pressure differences are relatively small, this assumption is valid. The assumption implies that the air density throughout the model is constant, thus $\rho = \rho_a$.

• The fill material is homogeneously distributed across the inlet section. There is no flow contraction and therefore, $A_{fr} = A_i$.

• The fill is the only flow obstruction in the tower. It is assumed that all the axial resistances in the vicinity of the fill are summed to arrive at the total resistance. Also, static flow obstructions before the fill (tower supports, inlet louvers) are not present in the model. Terblanche and Kröger [6] show that the inlet losses are independent of these obstructions in any case. Therefore, $K_t = K_{fi}$.

Employing these assumptions reduces Eq. (1) to,

$$K_{ct} = \frac{p_a - p_o}{\rho_a v_i^2/2} - 1 - K_{fi} \tag{2}$$

where $v_i = A_i/(\rho_a m_{av})$ and p_o is the static pressure after the fill. Equation (2) can now be applied to an experimental scale model.

The inlet losses for a circular cooling tower are dependent on the inlet diameter, d_i, the inlet height, H_i, and the radius of the inlet rounding, r_{ir}. The magnitude of the fill loss coefficient and other obstacles in its immediate vicinity, referred to as K_{fi}, also has a strong influence on the inlet loss. Applying dimensional analysis to the variables governing the inlet loss, reduces the functional dependency of the inlet loss coefficient to

$$K_{ct} = f\left(\frac{d_i}{H_i}, \frac{r_{ir}}{d_i}, K_{fi}\right) \tag{3}$$

By utilizing the description of dependencies provided above, the scale model shown in Fig. 2, that can vary each of the dimensionless groups independently (and thereby facilitate the formulation of empirical correlations), can be constructed.

The fills found in wet-cooling towers can be divided into two general categories: film type packing, such as fiber sheets, that offer an orthotropic resistance to airflow and splash type packing, such as trickle grids, that have an isotropic resistance (Fig. 3). These fill types cause different inlet losses when installed in geometrically similar cooling towers under identical conditions. The discrepancy is caused by the interaction of the recirculating air (Fig. 1) with the fill and by entrance effects when air enters the different fills at oblique angles. The result is that the inlet losses for isotropically and orthotropically packed towers must be determined separately.

A publication by Terblanche and Kröger [6] gives the inlet losses for counterflow dry-cooling towers fitted with radiator type heat exchangers. Since the geometry of film type fills and finned heat exchangers are similar the inlet losses predicted for the latter will also be valid for the former (neglecting the rain zone's influence). Terblanche's equations can therefore be used in conjunction with the present method.

For circular counterflow cooling towers with isotropic fills the experimental setup shown in Fig. 2 was used in conjunction with splash packing (Fig. 3) to determine the dry tower inlet losses. A curve fit of the results produced the following empirical correlation

$$\begin{aligned} K_{ct} = (&0.011266 e^{(0.093\, d_i/H_i)} K_{fi}^2 - 0.3105 e^{(0.1085\, d_i/H_i)} K_{fi} \\ &- 1.7522 + 4.5614 e^{(0.131\, d_i/H_i)}) \\ &+ \sinh^{-1}\left[\left(\frac{10970.2 e^{(-0.2442 K_{fi})} + 1391.3}{d_i/H_i - 15.7258} \right.\right. \\ &\left.\left. + 1205.54 e^{(-0.23 K_{fi})} + 109.314\right) \right. \\ &\left. \times \left(\frac{2 r_{ir}}{d_i} - \frac{0.01942}{d_i/H_i - 27.929} - 0.016866\right)\right] \end{aligned} \tag{4}$$

which is valid for $7.5 \leq d_i/H_i \leq 15$, $5 \leq K_{fi} \leq 25$ and $0 \leq r_{ir}/d_i \leq 0.02$. The average error encountered when fitting Eq. (4) to the experimental data is only 2.5 percent. The agreement is adequate for the purposes of this investigation, since experimental uncertainty can account for a deviation of up to 3 percent of the measured value.

The complexity of Eq. (4) is caused largely by the nonlinearity of the experimental data. The inverse hyperbolic sine function in the second term of Eq. (4) accounts for the inflection point in the r_{1r}/d_i dependent data shown in Fig. 4. Previous researchers neglected this dependency in favor of simplicity or used two different equations to predict the loss coefficient before and after the

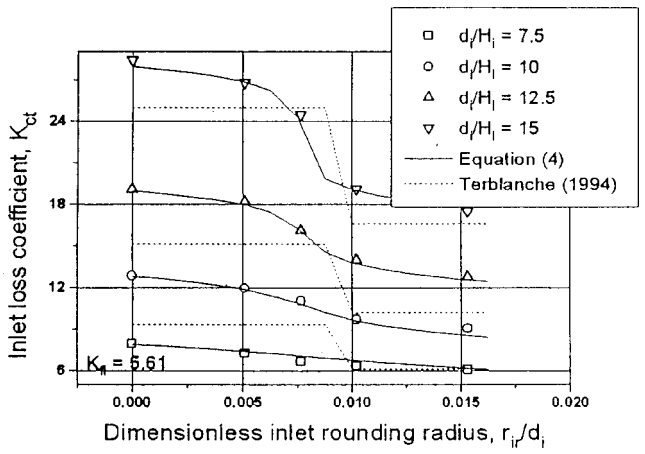


Fig. 4 Inlet loss coefficient for a circular tower

inflection point. The error in this approach seems to be significant, since the inlet loss coefficient continues to decrease, albeit slowly, with increased inlet rounding radius after the inflection point (Fig. 4).

A similar approach produced the following equation for the inlet loss coefficient in isotropically packed induced draft rectangular towers.

$$K_{ct}=0.2339+(3.919\times10^{-3}K_{fi}^2-6.840\times10^{-2}K_{fi}+2.5267)\times e^{W_i/H_i(0.5143-0.1803e^{(0.0163K_{fi})})}-\sinh^{-1}\left(2.77e^{(0.958\,W_i/H_i)}\times e^{(K_{fi}(2.457-1.015\,W_i/H_i)\times10^{-2})}\times\left(\frac{r_{ir}}{W_i}-0.013028\right)\right) \quad (5)$$

Figure 4 represents the curve fit of experimental data and clearly shows the difference in inlet loss between an isotropically packed tower (solid line) and an orthotropically packed tower (dotted line). The inlet loss associated with orthotropic resistance fills is generally smaller, especially at lower values of d_i/H_i.

Numerical Modeling of the Rain Zone

For the purposes of the numerical simulation, the commercial CFD code STAR-CD, was employed. Figure 5 depicts a typical computational mesh used to simulate a cooling tower inlet. The wedge shaped section represents a 10 degree arc of a circular cooling tower, similar to the experimental model shown in Fig. 2. Modeling the tower as a two-dimensional axisymmetrical problem significantly reduces computational time and geometrical complexity.

The mass flow rate at the outlet is considered fixed, while the inlet conditions are left unspecified, allowing a natural development of the flow field. Droplets were simulated using a time-derivative Lagrangian tracking scheme of sample drops and droplet deformation was taken into account according to a scheme proposed by Dreyer [8]. Turbulent effects were simulated by using the standard k-ε model, but the effects of the droplets and fill on turbulent characteristics were neglected because of insufficient data. The different fills were modeled as distributed resistances, the particulars of each type being incorporated in user-subroutines.

The pressure drop caused by the inlet losses, the rain zone, and the fill were measured by a row of specialized sensor cells inserted above the packing. The cells return the flow's piezometric pressure and vertical velocity, so that the total mechanical energy loss experienced by the flow, between ambient conditions and the top of the fill, can be calculated. The total loss coefficient, defined in terms of the average fill outlet velocity, is given by,

$$K_t=[p_a-\overline{((p_s+\tfrac{1}{2}\rho u_1^2)_{sc}+\rho g(H_i+L_{fi}))}]/(\tfrac{1}{2}\rho v_i^2) \quad (6)$$

where the subscript sc denotes conditions in the sensor cells, the overbar a mass flux based averaging process across the fill outlet,

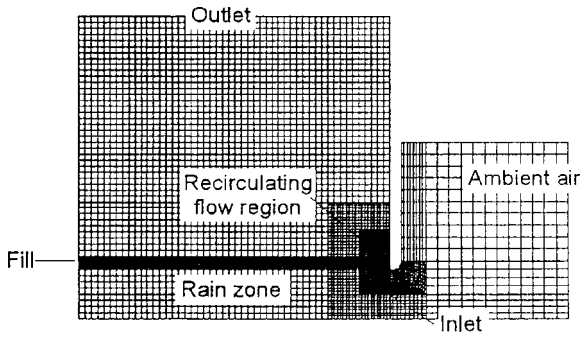


Fig. 5 Circular tower mesh with embedded refinement

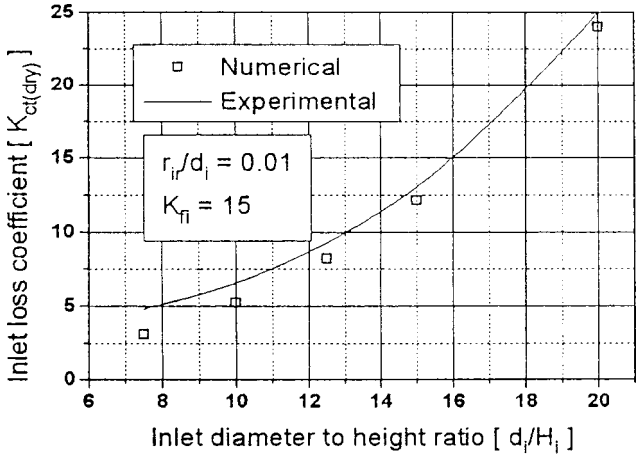


Fig. 6 Correspondence between experimental and numerical solutions for towers with isotropic fills

and p_a is the ambient stagnation pressure at ground level. The inlet loss coefficient for a wet-cooling tower can now be found from,

$$K_{ct}=K_t-(K_{fi}+K_{rz}) \quad (7)$$

where K_{rz} is the rain zone loss coefficient from De Villiers and Kröger [9].

As stated previously, inacurracies in the numerical turbulence model will cause a deviation from the actual physical value for the inlet loss coefficient (especially for isotropic fills). To compensate for this error, the experimental inlet loss coefficients for dry-cooling towers is modified in the following manner,

$$K_{ct(wet)}=K_{ct(dry)}\times dK_{ctrz} \quad (8)$$

where the inlet loss correction factor is

$$dK_{ctrz}=\frac{K_{ct(wet,numerical)}}{K_{ct(dry,numerical)}} \quad (9)$$

and $K_{ct(dry)}$ can be found from the experimental equations. This is an approximation of the wet-cooling tower inlet loss, but the size of the error will be bounded by the difference between the numerical and experimental dry-cooling tower values. It also serves to define trends in the interdependency of the inlet and rain zone losses.

The inlet loss correction factor, dK_{ctrz}, is a function of all eleven variables that characterize the inlet and rain zone losses. The time needed to explore all the dependencies and divergent influences of these variables using CFD is prohibitive. It was therefore decided to exclude weaker variables from the analysis after a through screening process. The ''strong'' functional dependency for the inlet loss correction factor is given by

$$dK_{ctrz}=f(d_i(\text{or } W_i),H_i,d_d,G_w,G_a,K_{fi}) \quad (10)$$

The general validity of the current analysis is dependent on the degree of similarity between the experimental and numerical simulations of the dry-cooling tower inlet loss coefficient. For a circular tower with an isotropic splash type fill, the correspondence between the numerical and experimental solutions can be seen in Fig. 6. The numerical solution's values are consistently lower than is the case for the experimental tests. This is in accordance with the predicted deficiency in the modeling of the recirculating flow region. In general, though, the agreement between the two approaches is sufficient to engender a large degree of confidence in further results obtained using the numerical model and the applicability of such results to the experimental results is assured.

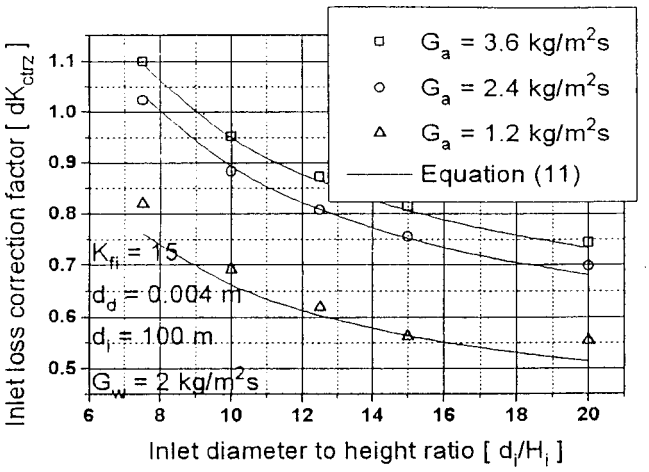


Fig. 7 Inlet loss correction factor for isotropically packed circular towers

For a circular counterflow wet-cooling tower, with an isotropic flow resistance fill, the correlation for the inlet loss correction factor is given by,

$$dK_{\text{ctrz}}=\left[0.2394+80.1\left(\frac{0.0954}{d_i/H_i}+d_d\right)e^{(0.395\,G_W/G_a)}-0.3195\left(\frac{G_W}{G_a}\right)-966.0\left(\frac{d_d}{d_i/H_i}\right)e^{(0.686\,G_w/G_a)}\right]\times(1-0.06825G_W)K_{\text{fi}}^{0.09667}e^{8.7434(1/d_i-0.01)} \quad (11)$$

which is valid for $7.5\leqslant d_i/H_i\leqslant 20$, $3\leqslant d_d\leqslant 6$ mm, $1\leqslant G_W\leqslant 3$ kg/m^{2}s, $1.2\leqslant G_a\leqslant 3.6$ kg/m^{2}s, $80\leqslant d_i\leqslant 120$ m, and $5\leqslant K_{\text{f1}}\leqslant 25$ (where $G_a=\rho_a v_i$ and $G_W=\rho_W v_W$). The lack of consistent use of dimensionless groups is caused by the exclusion of the ''weak'' variables, so that insufficient dimensional variables are available to complete the Π groupings. The average error encountered using the fitted correlation, is less than 4 percent, which is adequate considering the inaccuracies of the numerical simulation and the fact that the correction factor is used to modify an existing loss coefficient.

The general trend is for the added resistance of the rain zone to dampen the inlet loss in an effect similar to that caused by increased fill resistances as shown in Fig. 7. This change is not directly coupled to the rain zone loss, though, necessitating the large number of dependent variables.

The agreement between numerically and experimentally determined inlet losses for orthotropically packed circular dry-cooling towers is even better than for isotropic packing. The accuracy of the numerical solution is ascribed to the lack of large unbounded recirculating flows in the orthotropic film type fill. Upon investigation it was found that the inlet loss correction factor, for circular orthotropically packed towers, closely resembles that associated with isotropically packed towers, so much so that Eq. (11) represents an adequate correlation of the inlet loss correction factor for the orthotropic case.

The correspondence between experimentally and numerically generated inlet loss coefficients for isotropically packed rectangular towers is poor compared to that encountered in circular towers. This poor agreement is thought to be caused by a larger comparative recirculating region in rectangular towers, so that the error made during numerical modeling becomes correspondingly greater. Any results obtained from the CFD simulation must, therefore, be viewed with caution and lack the degree of confidence that is associated with the circular tower model. This is especially true at the low inlet width to height ratios with which operational rectangular towers are commonly constructed. Fortunately, the magnitude of the inlet loss in such towers is comparatively small and the inlet loss correction factor is close to unity, so that any expected improvement in tower performance, because of the correction factor, would be marginal. It therefore, becomes acceptable to ignore the inlet loss correction factor in cases where $W_i/H_i\leqslant 3$

Where the inlet loss correction factor for a rectangular tower is needed, the following empirical correlation provides the required value for both isotropic and orthotropic resistance fills,

$$dK_{\text{ctrz}}=1-G_w(0.123-12.1d_d-272.26d_d^2+5.04\times10^{-4}\cdot e^{0.466\,W_i/H_i})\times(1-1.16\times10^{-3}\cdot e^{G_a}) \quad (12)$$

and is valid for $3\leqslant W_i/H_i\leqslant 7.5$, $3\leqslant d_d\leqslant 6$ mm, $1\leqslant G_W\leqslant 3$ kg/m^{2}s and $2\leqslant G_a\leqslant 6$ kg/m^{2}s. It must be emphasized, that this equation can only be used with any degree of confidence at high W_i/H_i values and since this is not normally the case, it becomes prudent to take the conservative approach by ignoring the influence of the rain zone loss on the inlet loss.

Conclusion

Although cooling tower inlet losses have been extensively studied in the past, it was found that no correlations existed to predict the inlet loss for towers with isotropic resistance fills and/or rain zones. These deficiencies were addressed by, firstly, conducting experiments to find the inlet loss for dry isotropically packed towers. The data from these experiments was used to formulate correlations for the dry inlet loss coefficient in circular and rectangular towers. The results show that distinctly different inlet losses are encountered when employing isotropic resistance fills (compared to orthotropic resistances). In addition, increasing the inlet rounding's radius was found to continuously decrease the magnitude of the inlet loss, even after the inflection point reported by previous researchers.

Secondly, computational fluid dynamics (CFD) were used to investigate the effect of the rain zone's resistance on the inlet loss coefficient in circular and rectangular towers (with isotropic and orthotropic resistance fills). Equations for a correction factor were correlated from the results. Applying this correction factor to a dry inlet loss, produces an inlet loss (that is generally smaller than its dry counterpart) for an equivalent wet-cooling tower.

The correction factor is most applicable to circular towers, where good agreement, between experimental and numerical dry tower simulations, validates the modeling strategy. A ''weak'' agreement between numerical and experimental results for rectangular towers, caused by large recirculating regions and relatively small inlet losses, casts doubt on the efficacy of the corresponding correction factor. Studies of operational tower configurations show, however, that the influence of the correction factor in rectangular towers is small and may be ignored in favor of a more conservative approach. Whether the fill resistance is orthotropic or isotropic was found to have little effect on the size of the correction factor.

Overall, the prediction tools presented in this paper provide greater accuracy and confidence when calculating the mass flow rate and therefore the heat rejection rate in counterflow wet-cooling towers. It is suggested, however, that an attempt be made to validate the complete set of equations in terms of applicability to operational towers and effective droplet diameters.

Nomenclature

A = Area m^2
d = Diameter m
G = Mass velocity kg/sm^2
g = Gravitational acceleration m/s^2
H = Height m
K = Pressure drop coefficient
L = Length m
m = Mass flow rate kg/s

p = Pressure N/m^2
r = Radius m
v = Velocity m/s
W = Width m
z = Axial coordinate m

Greek symbols

ρ = Density kg/m^3

Subscripts

a = Air, Ambient
ct = Cooling tower
d = Droplet
fi = Fill, fill conditions
fr = Frontal
i = Inlet, into the fill
ir = Inlet rounding
L = Length
o = Outlet
r = Radial
rz = Rain zone
t = Total
v = Water vapor
w = Water
z = Axial

References

[1] Benocci, C., Buchlin, J. M., and Weinacht, P., 1986, ''Prediction of the Air-Droplet Interaction in the Inlet Section of a Natural Draft Cooling Tower,'' *Proceedings of the 5th IAHR Cooling Tower Workshop*, Monterey, CA.

[2] Lowe, H. J., and Christie, D. G., 1961, ''Heat Transfer and Pressure Drop Data on Cooling Tower Packing and Model Studies of the Resistance of Natural Draft Towers to Airflow,'' *International Development in Heat Transfer*, Part V, ASME, New York.

[3] Zembaty, W., and Konikowski, T., 1971, ''Untersuchung über den aerodynamischen Widerstand von Kühlturmen,'' Brennst. WarmeKraft, **23**, No. 10, pp. 441–445.

[4] Garde, M. A., 1981, ''Experimental Evaluation of Entrance Losses of Flared Natural Draft Cooling Towers,'' M.Sc. Thesis, Cornell University, Ithaca, NY.

[5] Geldenhuys, J. D., and Kröger, D. G., 1986, ''Aerodynamic Inlet Losses in Natural Draft Cooling Towers,'' *Proceedings of the 5th IAHR Cooling Tower Workshop*, Monterey, CA.

[6] Terblanche, J. E., and Kröger, D. G., 1994, ''Experimental Evaluation of the Aerodynamic Inlet Losses in Cooling Towers,'' South Afr. Inst. Mech. Eng. R & D J., **10**, No. 2, pp. 41–44.

[7] Du Preez, F., and Kröger, D. G., 1988, ''Experimental Evaluation of Aerodynamic Inlet Losses in Natural Draft Cooling Towers,'' *Proceedings of the 6th IAHR Cooling Tower Workshop*, Pisa, Italy.

[8] Dreyer, A. A., 1994, ''Modelling of Cooling Tower Splash Pack,'' Ph.D. dissertation, University of Stellenbosch, Stellenbosch.

[9] De Villiers, E., and Kröger, D. G., 1997, ''Analysis of Heat, Mass and Momentum Transfer in the Rain Zone of Counterflow Cooling Towers,'' *Proceedings of the 1997 IJPGC*, Vol. 2 PWR-Vol. 32, Denver, pp. 141–149.

Estimation of NO_x Emissions in Thermal Power Plants Using Neural Networks

G. Ferretti
e-mail: ferretti@elet.polimi.it

L. Piroddi
e-mail: piroddi@elet.polimi.it

Dipartimento di Elettronica e Informazione
Politecnico di Milano
Piazza Leonardo da Vinci 32,
Milano 20133, Italy

In this paper a neural network-based strategy is proposed for the estimation of the NO_x emissions in thermal power plants, fed with both oil and methane fuel. A detailed analysis based on a three-dimensional simulator of the combustion chamber has pointed out the local nature of the NO_x generation process, which takes place mainly in the burners' zones. This fact has been suitably exploited in developing a compound estimation procedure, which makes use of the trained neural network together with a classical one-dimensional model of the chamber. Two different learning procedures have been investigated, both based on the external inputs to the burners and a suitable mean cell temperature, while using local and global NO_x flow rates as learning signals, respectively. The approach has been assessed with respect to both simulated and experimental data. [DOI: 10.1115/1.1367339]

1 Introduction

Because of the harmful effects on environment and human health, nitrogen oxides (NO_x) emissions, produced by combustors and engines, have been the subject of restrictive regulations in many countries in recent years. As a consequence, both theoretical and experimental research on advanced abatement techniques have received an increasing interest. With reference to power plants, these techniques may be classified as *secondary* techniques, based on combustion gases treatment ([1]), and *primary* techniques, aimed at directly limiting NO_x generation through combustion control ([2]). The design of primary techniques, in particular, represents a complex control problem, where complexity stems from the thermodynamic and fluid-dynamic processes taking place in the combustion chamber and from the chemistry relevant to NO_x generation.

With respect to NO_x estimation, some significant results are available in the literatures, mainly for the case of gas turbines ([3–5]) and aviation and diesel fuel-air systems ([6–8]). These results are usually given in the form of correlations, obtained by means of a three-dimensional finite element approach ([4]), which is essential to derive accurate models of the overall combustion process. Though most of these correlations can be summarized by the following general form:

$$[NO_x] = f(P, \tau, T, \phi)$$

where P is the pressure, τ is the residence time, T is the ''temperature'' and ϕ is the fuel-air equivalence ratio, some significant differences among the various approaches are evident and no standard methodology has been yet established. For example, in Ref. [3] the equivalence ratio is replaced by the fraction of air in the primary zone combustion F, and T is defined as the ''stoichiometric'' temperature, while in [5] it is defined as the ''reaction'' temperature. Moreover, the residence time refers to the primary zone of the gas turbine, and it is in turn computed as a function of F and of the volume occupied in combustion V_c ([3]). Some modifications have been also introduced to include the effects of evaporation and mixing ([3,5]). The same correlations proposed in [5] are adopted in [9], with T computed this time as the ''flame'' temperature defined in [6] and giving ''reasonable'' values to τ.

In any case, the inputs configuration, the flow model and the operating conditions of combustors in gas turbines are very different from combustion chambers in boilers (not to mention the difference between the spatial scales). As a consequence, the application of the above correlations to the case of combustion chambers turns out to be difficult. With reference to power plants, the NO_x estimation problem must be therefore completely reformulated, and new correlation forms have to be investigated. In this respect, a three-dimensional approach seems to be essential for a thorough understanding of the physical processes involved, and also for a correct validation of any proposed estimation algorithm. However, from a computational point of view, it would be highly beneficial to adopt one-dimensional approaches, which are currently used, for control design and operators training, even in real-time mode.

A first attempt in this direction is discussed in [2], where a rather complex ad hoc one-dimensional model of a combustion chamber including NO_x generation is developed. The model is characterized by a separate modeling of different chamber zones, respectively, the *flame* and *gas* zone, and by few unknown parameters (mainly related to the modeling of turbulent diffusion), to be identified from experiments. Besides the model complexity, the NO_x estimation performance ranges from excellent to medium, depending on the operating conditions.

In this paper, a neural network approach to the prediction of NO_x emissions in thermal power plants is proposed. The approach is also based on a classical one-dimensional model of the combustion chamber, which plays a key role in describing, through the cell temperatures, the internal combustion conditions. In particular, one of the most critical tasks performed by the model is the computation of the gas emissivities and of the radiated heat.

In the development of the methodology, a three-dimensional finite element model of the combustion chamber has been used. One main result of the three-dimensional analysis is the evidence of the local nature of the oxide-generation process, bounded in the zones of the burners. Therefore, the overall NO_x production can be distributed among the three burner cells and described by the same local model. In view of this, a neural network NO_x estimator has been trained based on the cell temperatures and on the external inputs to the burners, i.e., fuel and air flow rates. Note that the approach allows different mixtures of oil and natural gas to be accounted for, as well as different oxide-generation mechanisms.

In a first version of the approach, called *Cell Learning*, a simple two-layered network structure and a standard backpropagation technique have been used, having the *local* NO_x production, com-

Contributed by the Power Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the Power Division, May 2000; final revision received by the ASME Headquarters January 2001. Editor: H. D. Nelson.

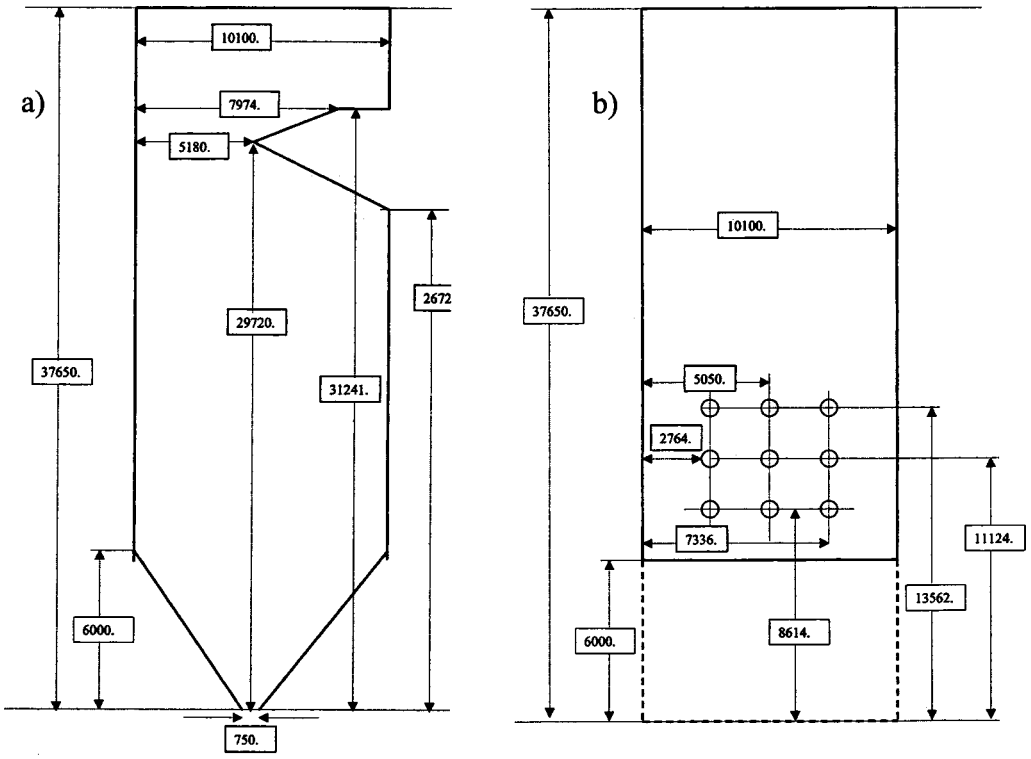


Fig. 1 Dimensions (mm) of the combustion chamber; (*a*) lateral view; (*b*) front view

puted by the three-dimensional model, as the teaching signal. The main purpose of this version was to validate the single cell model hypothesis and the role of the cell temperature. In fact, the Cell Learning implementation is prevented in practice, since the local NO_x production rates are normally not measured.

The *total* NO_x flow rate produced by the combustion chamber is indeed assumed as the teaching signal in the *Chamber Learning* approach, which maintains the paradigma of the single cell model by a suitable augmentation of the network structure and a simple modification of the backpropagation technique. The Chamber Learning has been successfully validated with both simulated and experimental data, and the robustness with respect to different data sets and operating conditions has been tested.

2 Three-Dimensional Model of the Combustion Chamber

As an essential aid to get a deep insight in the combustion process and, in particular, in pollutants generation, a three-dimensional finite volume model of the combustion chamber has been developed [10]. The model, based on the balance equations of mass, momentum, and energy, allows an accurate description of the thermo-fluid dynamics and of the main heat and mass transfer dynamics. The computations performed by the code are organized in four steps:

1 computation of the motion field (fluid dynamics);
2 computation of heat release (combustion);
3 computation of heat exchange and temperature field (heat transfer); and
4 computation of combustion products (chemical kinetics).

The furnace of a 320 MW unit of the power plant of Tavazzano (Italy), whose dimensions are reported in Fig. 1, is considered in this work. The furnace is endowed with nine burners in the front and nine burners in the back, placed on the walls at three levels in the lower part of the furnace (Fig. 1(*b*)), allowing combustion of both oil and methane. Recycling gases enter the furnace only at the bottom.

The volume of the combustion chamber has been discretized in 4059 cells, of dimensions $1.122\times0.918\times0.918$ m. This choice represents a tradeoff between accuracy and numerical efficiency. In fact, the three-dimensional model has been also developed to replace the one-dimensional model used in a real-time simulator of the plant, designed for training and control system design. As such, it had to comply with severe requirements in terms of computing time. In this way, some accuracy in describing the thermo-fluid dynamics over small spatial scales (cm) is lost.

The three-dimensional discretization of the combustion chamber was directly used to compute NO_x emissions, while an average over the same ten cells of the previous one-dimensional model was considered for the sake of thermal exchange and power generation modeling (thus allowing for a direct replacement of the one-dimensional model). Figure 2 shows the NO_x flow rate fields obtained from the three-dimensional model, assuming a 320 MW load and methane fuel. The fields are observed in two central orthogonal sections, respectively parallel (Fig. 2(*a*)) and orthogonal (Fig. 2(*b*)) to the planes of burners. Note that the fields are quite uniform along the horizontal coordinates and the generation of NO_x is limited to the sections containing the burners (3,4,5), located by the horizontal lines. The latter is particularly evident in Fig. 3, which shows the incremental generation of NO_x along the vertical coordinate.

This fact may be interpreted as follows. Generation of NO_x holds at the burners, where the highest temperatures are reached, by reaction of the nitrogen contained both in fuel (*fuel* NO_x) and in air (*thermal* and *prompt* NO_x). Since the temperature rapidly decreases when exiting from the zones of the burners, no new generation of NO_x takes place, while the reaction of NO_x destruction is equally inhibited. As an important consequence, the prediction of NO_x generation may be based on correlations relating

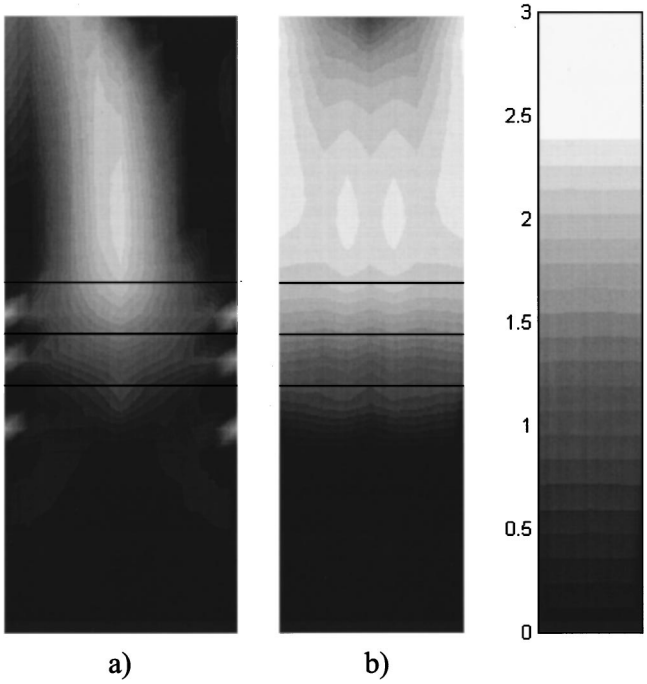


Fig. 2 NO_x flow rate fields (g/s)

the *local* (to the burners) NO_x flow rate to the external *local* inputs to the burners (oil, methane and air flow rates) and a suitable ''mean'' temperature.

At first, some classical correlations, similar to those proposed in Refs. [4], [5], [9], were tried out to predict NO_x generation, obtaining, however, poor results. It also became evident that the functional form of the correlations, developed and tested with reference to completely different combustion processes, was to blame for this outcome, rather than a wrong choice of the correlation variables.

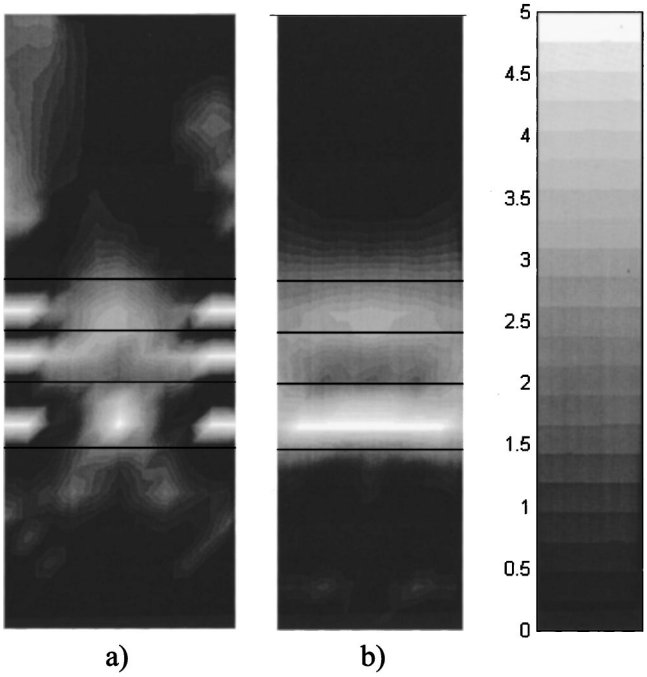


Fig. 3 Incremental NO_x flow rate fields (g/s)

To estimate a more adequate structure of the correlations a neural network approach was adopted, based exclusively on the external inputs of the burner cells (fuel and air inlet flow rates) and on an aggregate indicator of the state of combustion, such as the mean cell temperature or a ''flame'' temperature. It is important to note that the same emission model can be assumed for every burner cell. The overall NO_x emission can be computed by direct summation of the NO_x production rates in each burner cell, as estimated by the black-box models.

3 Estimation of NO_x Emissions Using Neural Networks

In order to evaluate the applicability of this approach, the three-dimensional model of the combustion chamber has been simulated in different conditions and a comprehensive set of data has been collected. Altogether, 35 simulation experiments were performed, ranging from a load of 140 MW to a load of 320 MW:

- 11 experiments at different load levels with methane fuel only;
- 14 experiments at different load levels with oil fuel only; and
- 10 experiments in various mixed-fuel combustion situations.

From each simulation experiment only data relative to the burner cells have been considered: the fuel $w_{g,i}$ (gas), $w_{o,i}$ (oil), and air $w_{a,i}$ inlet flow rates, the local NO_x production rate $w_{NO_x,i}$, and an aggregate cell temperature T_i. The index $i(i=1,2,3)$ refers to the specific burner cell. The overall NO_x production rate $w_{NO_x,out}$ is assumed equal to the sum of the three local production terms in the burner cells.

As already mentioned in the Introduction, the computation of the ''temperature'' is considered in the literature a critical step for the purpose of NO_x estimation, and various definitions of temperature are adopted. With specific reference to combustion chambers, the following analysis shows that while a cell temperature variable is essential for a consistent estimation of the NO_x emissions, the specific mode in which the aggregate cell temperature is computed is not critical.

Three different mean temperatures have been considered:

- the exponential mean cell temperature computed from the three-dimensional model;
- the linear mean cell temperature computed from the three-dimensional model; and
- the cell temperature computed from the one-dimensional model.

While the meaning of the latter two temperatures is clear, some comments are needed to introduce the exponential mean temperature. An exponential dependence of the NO_x generation of the form

$$[NO_x] = f(\cdot)e^{-k/T}$$

with $f(\cdot)$ being some function of inputs flow rates, mass fractions, etc., where the importance of an accurate estimation of temperature peaks is evident, appears in several correlations as well as in the Zeld'ovich model ([3,5,11,12]). For this reason, our first guess was to weigh exponentially the three-dimensional temperature cell profile, according to the following formula:

$$\tilde{T} = -\frac{k}{\ln\left(\frac{1}{V}\int_V e^{-k/T}dV\right)}$$

where a value of 70,000 has been assumed for k (chosen as an average of the values used in the cited papers).

A conventional feed-forward neural network structure, with a single hidden layer with sigmoidal nonlinearities and a linear output layer, has been assumed for the black-box model. Two different learning procedures have been developed:

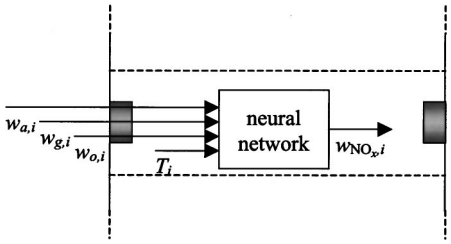


Fig. 4 Neural network structure for cell learning

• *Cell Learning*

The complete data set is used in the training phase and the *local* NO_x production rate constitutes the teaching signal. A standard backpropagation technique can be employed in this case. Cell Learning is useful to validate the single cell model hypothesis and to evaluate the role of the temperature variable. However, it can be applied only to simulated data, since normally the local NO_x production rates are not measured.

• *Chamber Learning*

The *total* NO_x emission is assumed as the teaching signal. Anyway, as explained in the appropriate section, it is still possible to maintain the paradigma of the single cell model by means of a suitable modification of the backpropagation technique. Chamber Learning is the only technique applicable in practice, when the local NO_x production rates are not known. However, it can profit greatly from a previous stage of Cell Learning based on accurate simulation data.

3.1 Cell Learning. A sketch of the neural network input/output structure is outlined in the following Fig. 4. The inputs are the air ($w_{a,i}$), methane ($w_{g,i}$), and oil ($w_{o,i}$) flow rates in the ith burner cell and the corresponding cell temperature (T_i). The output is the estimated local NO_x production rate ($w_{NO_x,i}$). All the input/output data are considered equivalently, no matter what burner cell they pertain to. A standard backpropagation technique has been employed to train the feedforward neural network parameters. With a trial and error procedure a hidden layer of 15 neurons has been selected.

Three data sets have been employed to test the effect of using the three different cell temperature computation techniques above defined. The estimation performance of the trained neural network on the data set referring to the one-dimensional temperature is reported in Fig. 5. Similar results are obtained for the other data sets. The continuous line represents the true (simulated) local NO_x values, while the dotted line is the corresponding neural estimate. The 105 data are reported consecutively, e.g., the data labeled $3x-2$, to $3x$, represent the values associated to the three burner cells in simulation No. x (simulations 1 to 11 are with methane fuel only, simulations 12 to 25 are with oil fuel only and simulations 26 to 35 are with mixed fuel). A performance comparison of the neural networks estimated on the basis of the three different temperatures is given in Table 1 in terms of various estimation error indices.

These simulation experiments display an almost equivalent and satisfactory estimation performance, demonstrating how even the one-dimensional model temperatures can be used effectively for the estimation of local NO_x emissions. On the other hand, there is no need to impose an exponential dependency on the temperature directly at the network input: It is the network training that selects the appropriate function dependency.

The reader should not be misled to think that the role of the temperature variable is inessential for the NO_x estimation: In fact, many of the data considered have the same fuel and air flow rates in every burner cell, while different NO_x flow rates are experienced in every cell, due to the different combustion conditions.

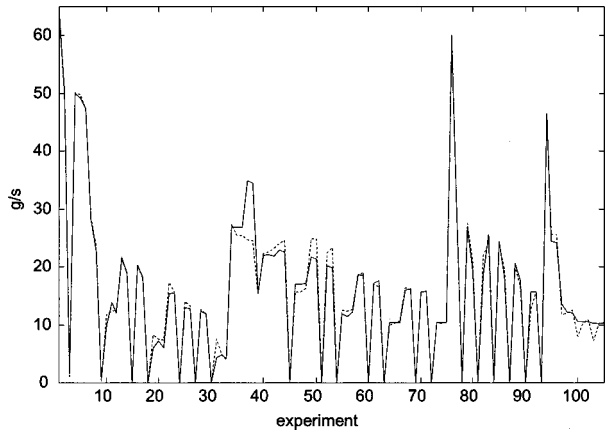


Fig. 5 Cell Learning with one-dimensional model cell temperatures

Thus, discarding the temperature information would prevent the neural network model from discriminating between them.

3.2 Chamber Learning. In the Chamber Learning mode, a virtual three-layered network must be considered, as in Fig. 6. In fact, the specific network structure, which contains three not interconnected replicas of the same smaller network, can be exploited to simplify the training algorithm. Consider the parameter adaptation laws for the Cell Learning case. The neural network is composed of two layers, the nonlinear hidden layer (index 1) and a linear output layer (index 2):

$$h^i_{1,j}=\sum_{k=1}^{N_k} w_{1,jk}u^i_k$$

$$y^i_{1,j}=g_1(h^i_{1,j})$$

Table 1 Estimation error indices with different mean temperature

Cell Learning	max(|err|)	mean(|err|)	MSQE
3D-simulator mean exponential temperatures	11.8645	0.9899	3.6143
3D-simulator linear mean temperatures	12.8124	0.9988	3.3154
1D-model cell temperatures	10.2215	0.9676	3.3036

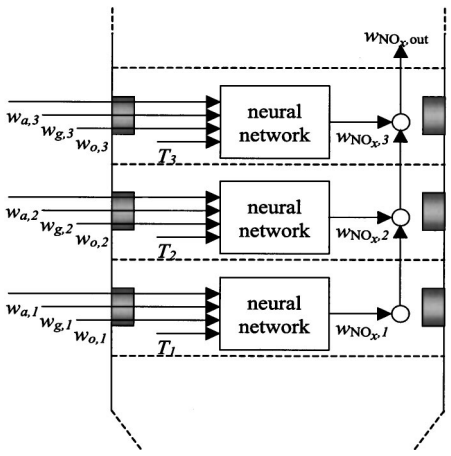


Fig. 6 Neural network structure for chamber learning

$$y_2^i = h_2^i = \sum_{j=1}^{N_j} w_{2,j} y_{1,j}^i$$

($i=1,2,3$, $N_k=4$, $N_j=15$) where $w_{1,jk}$ and $w_{2,j}$ are the network weights, u_k^i represent the network inputs, $y_{1,j}^i$ and y_2^i are the outputs of the two layers (y_2^i is also the whole network's output, i.e., the estimate of the NO_x flow rate of the ith burner cell), $h_{1,j}^i$ and h_2^i denote the results of the weighted linear combination performed in each layer, and $g_1(\cdot)$ is a nonlinear sigmoidal function. The following MSQE is minimized with respect to the network weights:

$$E(w) = \frac{1}{2} \sum_P \sum_{i=1}^{3} (\overline{y_2^i} - y_2^i)^2 = \frac{1}{2} \sum_P \sum_{i=1}^{3} (e_2^i)^2$$

where $\overline{y_2^i}$ represents the actual output value and e_2^i represents the cell NO_x flow rate estimation error. The sum of the squared estimation errors is extended over all available training patterns (P denotes the training set of 35 simulation experiments). The network weights are iteratively changed based on the following formulas:

$$\Delta w_{2,j} = -\eta \frac{\partial E}{\partial w_{2,j}} = \eta \sum_P \sum_{i=1}^{3} e_2^i y_{1,j}^i$$

$$\Delta w_{1,jk} = -\eta \frac{\partial E}{\partial w_{1,jk}} = \eta \sum_P \sum_{i=1}^{3} e_2^i w_{2,j} g_1'(h_{1,j}^i) u_k^i .$$

In the Chamber Learning case the neural network has a third linear layer composed by a single neuron with fixed unitary weights, which performs the sum of the outputs of three identical instances of the Cell Learning neural network, corresponding to the three burner cells:

$$y_3 = \sum_{i=1}^{3} y_2^i$$

where the index i ranging from 1 to 3 denotes a specific burner cell. The error function is different from the previous case, since the total NO_x at the output of the combustion chamber must be used as the training signal ($\overline{y_3}$):

$$E(w) = \frac{1}{2} \sum_P (\overline{y_3} - y_3)^2 = \frac{1}{2} \sum_P e_3^2 .$$

The resulting back propagation learning rules are reported below. The $w_{2,j}$ parameter variation can be computed as the sum of the individual variations of the same parameter computed for each burner cell network as a result of the specific cell inputs and an e_2^i error equal to e_3.

$$\Delta w_{2,j} = -\eta \frac{\partial E}{\partial w_{2,j}} = \eta \sum_P e_3 \sum_{i=1}^{3} y_{1,j}^i = \eta \sum_P \sum_{i=1}^{3} e_3 y_{1,j}^i$$

The $w_{1,jk}$ parameter variation can be computed as the sum of the individual variations of the same parameter computed for each burner cell network as a result of the specific cell inputs and an e_2^i error equal to e_3.

$$\Delta w_{1,jk} = -\eta \frac{\partial E}{\partial w_{1,jk}} = -\eta \sum_{i=1}^{3} \frac{\partial E}{\partial y_{1,j}^i} \frac{\partial y_{1,j}^i}{\partial w_{1,jk}}$$

$$= \eta \sum_P \sum_{i=1}^{3} e_3 w_{2,j} g_1'(h_{1,j}^i) u_k^i$$

Comparing these weight adaptation laws with those obtained in the previous case, the resulting algorithm can be seen to be equivalent to the standard backpropagation algorithm applied to the single cell network as in the Cell Learning mode, with the only difference that the local NO_x emission estimation error must be substituted with the global NO_x emission estimation error of the corresponding simulation.

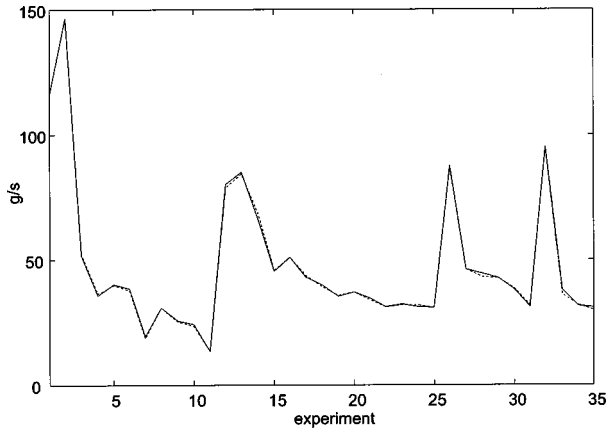


Fig. 7 Chamber learning estimation of the global NO_x emissions

At a cost of a worsened performance in the local cell NO_x estimation with respect to the Cell Learning results, this method achieves an extraordinary quality of estimation of the global NO_x emissions (see Fig. 7, where the three-dimensional simulator linear mean temperatures have been employed). The corresponding MSQE is as low as 0.5287. The same estimation experiment has been reproduced using *measured* global NO_x emissions obtained in nine experiments at different load levels with methane fuel only. As it is apparent from Fig. 8 the performance is comparable to that obtained with simulation data, and represents a further validation of the proposed approach.

3.3 Robustness Analysis. The obtained estimation results must be validated to ensure that the proposed approach is robust and generalizable to different operating conditions. In fact, if the neural network is overtrained, it will not be capable of describing new data, not previously included in the training set, unless the training phase is restarted. On the other hand, if a robust estimation technique is adopted, the neural network may be able to generalize to new data. A classical approach consists in preventing the network from overfitting the data, by stopping the training phase when the network performs satisfactorily on completely independent data.

To this aim the available data set has been split in a training set (30 simulations) and a validation set (5 simulations). Training has

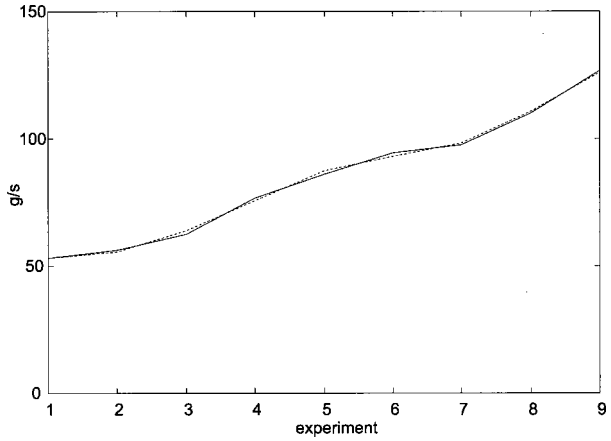


Fig. 8 Chamber learning estimation of the global NO_x emissions with measured data

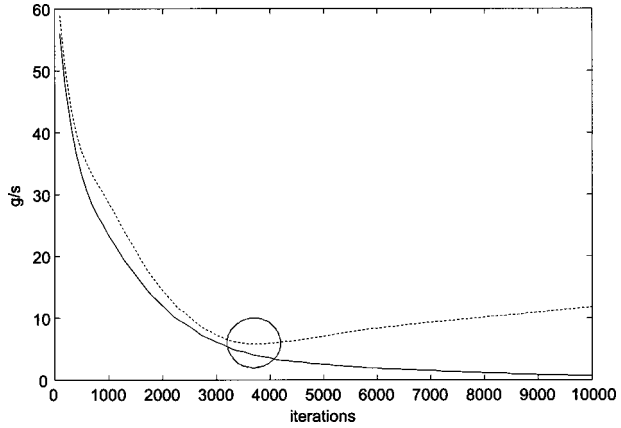


Fig. 9 Evolution of the MSQE during the training phase (continuous line=training set, dotted line=validation set)

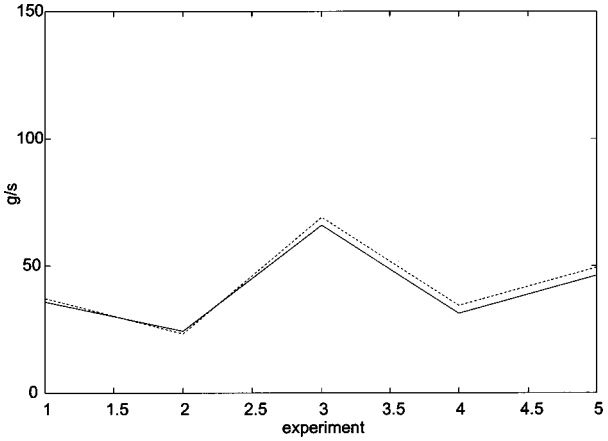


Fig. 11 Estimation performance on the validation set

been performed on the training set, while monitoring the performance on the validation set, and the estimation algorithm was stopped when a minimum MSQE was obtained on the latter data. The evolution of the MSQE during the training phase is reported in Fig. 9. As a result of the anticipated stopping of the learning algorithm, partially worse results are to be expected, as documented in Table 2.

However, the performance on both data sets remains acceptable (see Figs. 10–11).

4 Conclusions

A novel approach for the estimation of NO_x emissions in thermal power plants has been proposed. The approach has been developed with the essential aid of a three-dimensional model of the combustion chamber, which allowed to ascertain the local nature of the NO_x generation process, bounded to the zones of the burners. As a main consequence of this fact, the estimation model relies only on the external inputs to the burners (fuel and air flow rates), and on a single local internal variable: the mean cell temperature computed by a classical one-dimensional model of the chamber. These variables are used by a feedforward neural network to estimate the NO_x emissions, according to two different learning procedures, named Cell and Chamber Learning, respectively. Cell Learning uses the local NO_x flow rate as the teaching signal: as such it is not applicable in practice, when only the global NO_x flow rate is measured, and is also a computationally expensive method, since it requires the use of the three-dimensional simulator. However, it proved to be an essential step in the development of the Chamber Learning method, which does not require knowledge of the local NO_x emissions. For this reason, Chamber Learning lends itself to a direct implementation on the real system. Furthermore, it requires the use of a classical one-dimensional model only, instead of an expensive three-dimensional simulator. Another feature of the proposed method is the ability to deal with different mixtures of oil and methane fuels.

The approach has been validated with both simulated and experimental data, and a neural network robustness analysis has also been performed. It is expected that the proposed approach can be easily adapted to different types of combustion chambers.

The use of the NO_x estimation algorithm is currently under study with reference to a primary abatement technique, mainly based on a correct mixing of fuels.

Table 2 Estimation error indices with the training and validation set

Data set	max(|err|)	mean(|err|)	MSQE
Training set	5.0229	1.6825	3.8819
Validation set	3.0331	2.2364	5.7927

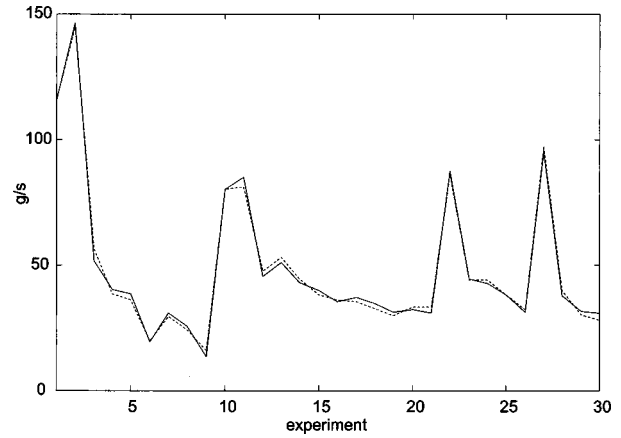


Fig. 10 Estimation performance on the training set

Acknowledgments

The authors would like to thank F. Pretolani and E. Tassi of CESI, for their valuable suggestions and experimental support.

This work has been supported by CESI.

References

[1] Belli, P., Bittanti, S., Campi, M., De Marco, A., Ferretti, A., Malloggi, S., and Prandoni, W., 1996, ''A Control System for Nitrogen Oxides Pollution Abatement by SCR (Selective Catalytic Reduction),'' *13th IFAC Triennal World Congress*, Vol. O, San Francisco, pp. 13–18.

[2] De Marco, A., and Poncia, G., 1999, ''A Model of Combustion Chambers, Including Nitrogen Oxide Generation, in Thermal Power Plants,'' Control Engineering Practice, **7**, pp. 483–492.

[3] Lefebvre, A. H., 1984, ''Fuel Effects on Gas Turbine Combustion-Liner Temperature, Pattern Factor, and Pollutant Emissions,'' J. Aircr., **21**, No. 11, pp. 887–898.

[4] Rizk, N. K., and Mongia, H. C., 1993, ''Three-Dimensional Gas Turbine Combustor Emissions Modeling,'' ASME J. Eng. Gas Turbines Power, **115**, pp. 603–611.

[5] Rizk, N. K., and Mongia, H. C., 1993, ''Semianalytical Correlations for NO_x, CO and UHC Emissions,'' ASME J. Eng. Gas Turbines Power, **115**, pp. 612–619.

[6] Gülder, Ö. L., 1986, ''Flame Temperature Estimation of Conventional and Future Jet Fuels,'' ASME J. Eng. Gas Turbines Power, **108**, pp. 376–380.

[7] Gülder, Ö. L., 1986, ''Combustion Gas Properties—Part II: Prediction of Partial Pressures of CO_2 and H_2O in Combustion Gases of Aviation and Diesel Fuels,'' ASME J. Eng. Gas Turbines Power, **108**, pp. 455–459.

[8] Gülder, Ö. L., 1988, ''Combustion Gas Properties—Part III: Prediction of the Thermodynamic Properties of Combustion Gases of Aviation and Diesel Fuels,'' ASME J. Eng. Gas Turbines Power, **110**, pp. 94–99.
[9] Fattori, M., Pittaluga, F., Beltrami, F., and Torelli, A., 1996, ''Theoretical Predictions of Thermochemical Performances and Pollutant Emissions of Heavy-Duty Gas Turbine Combustors by Means of Extended-Parametrization Modeling,'' TURBOMACCHINE'96, pp. 121–132.
[10] Castellano, L., 1998, ''Manuale a corredo della relazione tecnica finale del simulatore della caldaia TAVAZZANO UP Gruppo 6,'' Technical Report Matec MTC/FL143/98, Milano, Italy (in Italian).
[11] Makansi, J., 1988, ''Reducing NO_x Emissions,'' Power, pp. S1–S13.
[12] Warnatz, U., Maas, R., and Dibble, W., 1996, *Combustion: Physical and Chemical Fundamentals, Modelling and Simulation, Experiments, Pollutant Formation*, Springer, Berlin.

Journal of Engineering for Gas Turbines and Power

Technical Brief

A Theoretical Model of Uniform Flow Distribution for the Admission of High-Energy Fluids to a Surface Steam Condenser

J. Wang[1]
e-mail: j-wang@sheffield.ac.uk

G. H. Priestman

Department of Chemical and Process Engineering,
The University of Sheffield, Sheffield, S1 3JD, England

D. Wu
School of Mechanical Engineering, East China University of Science and Technology, Shanghai 200237, P.R. China

An analytical study is made of the perforated pipe distributor for the admission of high-energy fluids to a surface steam condenser. The results show that for all perforated pipes there is a general characteristic parameter M (kD/Lf), which depends on the pipe geometry and flow properties. Four cases are considered based on the value of the characteristic parameter M. (1) When M $\geqslant 1/4$*, momentum controls and the main channel static pressure will increase in the direction of the streamline. (2) When* $1/6 \leqslant M < 1/4$*, the momentum effect balances friction losses and the pressure will decrease to a minimum, and then increase in the direction of flow to a positive value. (3) When* $0 < M < 1/6$*, friction controls and the pressure will decrease to a minimum, then increase slowly, but the total pipe static pressure difference will always be negative. (4) When* $M = 0$*, a limiting case when the ratio of the length to the diameter is infinite. This analysis is useful not only for the design of perforated pipe distributors for turbine condensers over a wide range of dimensions, fluid properties, and side hole pressure but also for many other technical systems requiring branching flow distribution.*
[DOI: 10.1115/1.1359237]

Introduction

Accompanying the development of power generating technology is an increase in the fluid state properties in boilers and steam turbines. The amount, volume, and level of high-energy fluids, such as exhaust steam and drains from the by-pass or boiler, are also increased during the unit start-up, shut-down, or load decreases. The high energy exhaust fluids are received by the turbine condenser, and could exceed the design heat load of condenser. This can affect normal working, resulting in erosion of metallic parts or baffles, vibration, or even thermal deformation. Although the turbine condensers have been commercially available for many years, the exhaust equipment represents a problem that is often encountered in the turbine condenser. One of the many types of exhaust equipment consists of a main cylindrical channel on which a row of small holes are drilled along its length. Such exhaust equipment is called a perforated pipe.

The fluid entering the perforated pipe is divided into several branching streams, and then expanded into the condenser. Hence, the design of the perforated pipe in the condenser includes two steps: (1) uniform flow distribution; (2) parallel multiple jet expansion ([1,2]). The perforated pipe was recently studied in the literature ([1–3]). Sebald et al. studied the pressure drop and flow distribution applying the Bernoulli theorem. To date, no new effort has been directed towards this area, and only minimal understanding of the mechanisms have been developed. Thus data is particularly lacking on flow characteristics for such exhaust perforated pipes in turbine condensers. Most of the available data deals with pressure drop and flow characteristics under the assumption of uniform flow distribution and has been published in other branching flow systems by Bassiouny and Martin [4], Chang et al. [5], and Jin et al. [6].

This paper presents an attempt to introduce a general theoretical method to calculate flow distribution and pressure drop in the perforated pipe, where the aim is to provide for equal flow rates through the perforated pipe, and not to include multiple-jet expansion. The results are not limited to application on turbine condenser perforated pipes, but can be also applied to similar branching problems found in plate heat exchangers, air-conditioning installations, fluidized beds, gas burners, and heating systems.

Theoretical Model and Analyses

The dividing flow device often used for the turbine condenser can be represented schematically as shown in Fig. 1. It is constructed from a main channel of constant cross section that terminates in a closed end and is provided with equally spaced, uniformly sized side holes which are perpendicular to the main channel.

The variation of fluid pressure arises from two causes: (a) the friction of the fluid against the internal surface of the main channel makes the pressure fall in the direction of flow; and (b) the momentum of the main fluid stream flowing into a hole tends to carry the fluid toward the closed end, where an excess pressure is produced due to the deceleration of the flow. When the large fluid stream flows into the opening hole and undergoes subdivision the friction and momentum effects work in opposite directions, the first tending to produce a pressure drop and the second a pressure rise.

[1]To whom correspondence should be addressed.

Contributed by the Power Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the Power Division, Sept. 2000; final revision received by the ASME Headquarters Jan. 23, 2001. Editor: D. Lou.

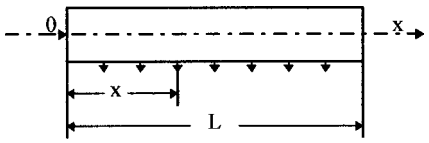


Fig. 1 Perforated pipe distributor

The development of the theoretical flow model is based on the following suppositions: (a) the exhaust pressure of the holes is constant and equal to the condenser's pressure; (b) a main channel has constant cross section and is horizontal; (c) the side holes are equally spaced, uniformly sized, and at right angles to the main channel; (d) the flow velocity of the main channel is highest at the open end and zero at the closed end.

Consider a section of the main channel near one of the branching outlets as shown in Fig. 2. The hole port can be regarded for all practical cases as a quasi-continuous system. Therefore the mass and momentum balance on the control volume may be written in the following way:

(1) Mass conservation

$$\rho F W = \rho F\left(W + \frac{dW}{dx}dx\right) + \rho F_c U_c$$

The basic equation becomes

$$U_c = -\frac{FL}{F_c n}\frac{dW}{dx} \tag{1}$$

where $dx = L/n$

(2) Momentum conservation

All axial components of velocity change result from the combined effects in axial pressure and friction force:

$$PF - \left(P + \frac{dP}{dx}dx\right)F - \tau_W \pi \; dDX = \rho F\left(W + \frac{dW}{dx}dx\right)^2 - \rho F W^2 + \rho F_c U_c W_c$$

The simplest shape of main channel is a circular cross section in which its friction force $\tau_w = \rho F(W^2/8)$. Neglecting the higher orders of dx and collecting the terms, one obtains:

$$\frac{1}{\rho}\frac{dP}{dx} + \frac{f}{2D}W^2 + 2W\frac{dW}{dx} + \frac{F_c n}{FL}U_c W_c = 0 \tag{2}$$

where W_c is the axial velocity component of the fluid in the pipe, which will be branched off through the hole. It can be smaller or larger than W, depending on the dimensions of the pipe and location of the holes. It may be then expressed as a fraction of W as given by Bassiouny and Martin [3] as follows:

$$W_c = \beta W \tag{3}$$

After substituting Eqs. (1) and (3) into Eq. (2), the momentum equation may be given as

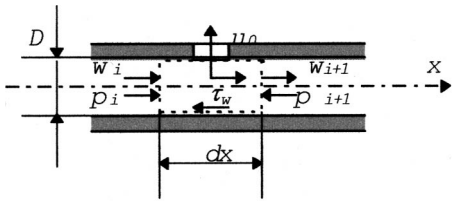


Fig. 2 Control volume near a hole

$$\frac{1}{\rho}\frac{dP}{dx} + \frac{f}{2D}W^2 + (2-\beta)W\frac{dW}{dx} = 0 \tag{4}$$

After substituting $2k$ for $2-\beta$ Eq. (4) becomes formally as given in references ([2,5]) as follows:

$$\frac{1}{\rho}\frac{dP}{dx} + \frac{f}{2D}W^2 + 2kW\frac{dW}{dx} = 0 \tag{5}$$

In general, k is expected to be smaller than unity.

Since the holes are the same size and their distribution along the length of the pipe is constant, uniform fluid distribution requires that the rate of flow through the pipe must vary linearly from a maximum at the inlet to zero at the closed end, or

$$W_X = W_0\left(1 - \frac{x}{L}\right) \tag{6}$$

Inserting Eq. (6) into (5) yields an ordinary differential equation for the pressure in the pipe which can be integrated from zero to x. Finally, one obtains

$$\Delta P_{x0} = \rho W_0^2 k\left[1 - \left(1 - \frac{x}{L}\right)^2\right] - \frac{\rho W_0^2 L f}{6D}\left[1 - \left(1 - \frac{x}{L}\right)^3\right] \tag{7}$$

After introducing the following dimensionless groups:

$$\overline{P} = \frac{P}{\rho W_0^2} \text{ and } \overline{X} = \frac{x}{L}$$

Eq. (7) can be given as follows:

$$\Delta \overline{P}_{X0} = k[1 - (1-\overline{X})^2] - \frac{Lf}{6D}[1 - (1-\overline{X})^3] \tag{8}$$

Differentiating Eq. (8) we receive the first derivative and the second derivative:

$$\Delta \overline{P}'_{x0} = (1-\overline{X})\left[2k - \frac{Lf}{2D}(1-\overline{X})\right] \tag{9}$$

$$\Delta \overline{P}''_{x0} = \frac{Lf}{D}(1-\overline{X}) - 2k \tag{10}$$

Let $\Delta \overline{P}'_{X0} = 0$ the function $\Delta \overline{P}_{x0}$ has two points of extreme values:

$$\overline{X}_{c1} = 1, \quad \overline{X}_{c2} = 1 - \frac{4kD}{Lf} \tag{11}$$

where $M = kD/L\underline{f}$.

Inserting $\overline{X}_{c1}, \overline{X}_{c2}$ into Eq. (10) yields:

$\Delta \overline{P}''_{x0}|_{\overline{X}_{c1}=1} = -2k < 0$, *i.e.* $\overline{X}_{c1} = 1$ is a point of maximum.

$\Delta \overline{P}''_{x0}|_{\overline{X}_{c2}=1-4kD/Lf} = 2k > 0$, *i.e.* $\overline{X}_{c2} = 1 - \dfrac{4kD}{Lf}$ is a point of minimum.

There is a maximum at the point $\overline{X}_{c1} = 1$, *i.e.*

$$\Delta \overline{P}_{X0}|_{\overline{X}_{c1}} = k - \frac{Lf}{6D} \tag{12}$$

and a minimum at the point $\overline{X}_{c2} = 1 - 4kD/Lf$, *i.e.*

$$\Delta \overline{P}_{x0}|_{\overline{X}_{c2}} = k\left[1 - \left(\frac{4kD}{Lf}\right)^2\right] - \frac{Lf}{6D}\left[1 - \left(\frac{4kD}{Lf}\right)^3\right] \tag{13}$$

Results and Discussion

The division of a fluid stream into parts by means of a perforated pipe distributor is accompanied by fluid pressure changes owing to wall friction and to the changing fluid momentum. Fric-

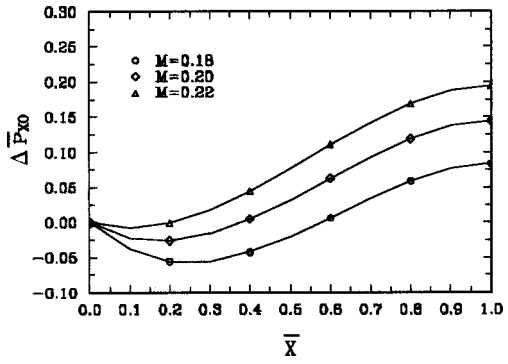


Fig. 3 The static pressure profile of the main channel for $1/6 \leqslant M < 1/4$

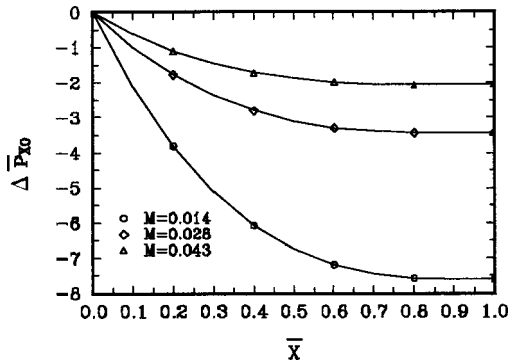


Fig. 5 The static pressure profile of the main channel for $0 < M < 1/6$

tion tends to the make the pressure fall, and momentum exchange tends to make the pressure rise. When these pressure trends are exactly balanced, the static pressure remains constant along the entire length of the perforated pipe, which insures a uniform flow distribution. We easily determine the greatest value and the smallest value of this function in a closed interval [0, 1] by comparing the extreme values of a function $\Delta \bar{P}_{x0}$ with the end-point values. The greatest (smallest) one of the absolute value of these values will be the greatest deviation of flow distribution. Although it is not possible to keep the fluid pressure perfectly uniform inside the main channel, based on Eq. (8) and Fig. 3, it is possible that the pressure variation can be limited to a minimum amount that produces a small but tolerable inequality in flow through the holes.

$\bar{X}_{C2} = 1 - 4M$ indicates that $\bar{X}_{c2}$ depends on the value M, i.e. the extreme values of Eq. (8) will vary as M. The treatment of Eq. (8) gives four different cases.

Case 1: $M \geqslant 1/4$ ($\bar{X}_{c2} = 1 - 4M \leqslant 0$), Momentum Controls. In this case, there is no minimum value in the interval [0, 1] because of $\bar{X}_{c2} < 0$. The main fluid stream is decelerated due to the loss of fluid through the lateral perforations. The momentum exchange has exceeded the friction loss since the fluid had entered the open end. Therefore, the static pressure will monotonically rise in the flow direction, and reach the maximum at the closed end. On the other hand, the static pressure differences increase as M increases. Figure 4 shows a typical set of pressure curves to this case. This case corresponds to a larger k or D/L, and a smaller f, and often occurs in chemical processing systems.

Case 2: $1/6 \leqslant M < 1/4$ ($1/3 \geqslant \bar{X}_{c2} = 1 - 4M > 0$), Momentum Effect Balances Friction. Owing to the fact that the friction and momentum effects work in opposite directions, the former tends to produce a pressure drop and the latter a pressure rise. There is a minimum at the point $\bar{X}_{c2}$. When $\bar{X} < \bar{X}_{c2}$ the friction effect exceeds the momentum, which makes the pressure fall in the direction of flow; when $\bar{X} > \bar{X}_{c2}$ the momentum effect increases slowly, and the pressure rises. Finally, the momentum effect will balance or exceed the friction, and the static pressure varies from negative to positive in the flow direction as shown in Fig. 3. It is apparent that the function $\Delta \bar{P}_{x0}$ could not be a constant for any value M. That is to say, a perfectly uniform side flow is impossible. If there is a smaller deviation for the side flow, this case may be $M = 1/6$, according to Eq. (8) or Fig. 4.

Case 3: $0 < M < 1/6$ ($1 > \bar{X}_{c2} = 1 - 4M > 1/3$), Fiction Controls. The opposite of case 1, the momentum effect is smaller than the friction, and the pressure difference is always negative along the flow direction. Although there is a very small pressure rise when $\bar{X} > \bar{X}_{c2}$ in the closed interval [0, 1], it is too small to be apparent in Fig. 5. The greater M is, the smaller is the pressure fall.

If the values of f and k are fixed, M will decrease with increasing L/D and the pressure of the main channel will decrease. In unit operations, it is mainly limited to flow in tiny-long channels, the flow of some non-Newtonian fluids, and extremely viscous Newtonian fluids. Also, it can be seen in ventilating systems or irrigation systems.

Case 4: $M \leqslant 0$ ($\bar{X}_{c2} = 1 - 4M \geqslant 1$), a Limiting Case or Not Existant. There is no momentum effect in the channel for $M = 0$. In practice it denotes a limiting case, for example, the ratio L/D or f is of infinitely large magnitudes. This case may be seen in irrigation systems.

The case $M < 0$ should not exist when all of k, D, L and f are positive.

Finally, Fig. 6 represents the effect of changing the parameter

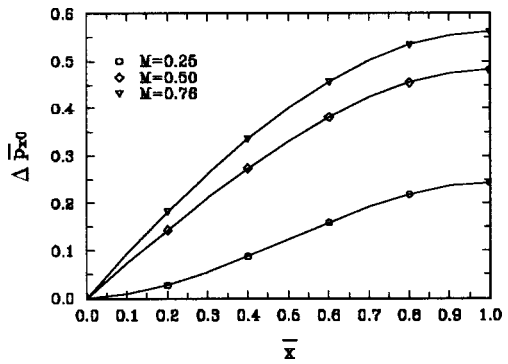


Fig. 4 The static pressure profile of the main channel for $M \geqslant 1/4$

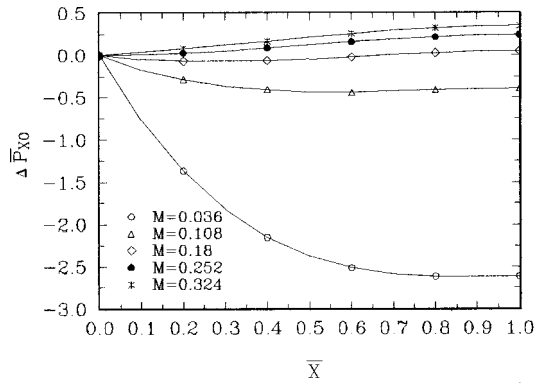


Fig. 6 The effect of changing M on flow pressure

M on the flow pressure (or the flow distribution) in the main channel. The effects of M on flow distribution is more apparent at a small M than at a large M.

The advantage of the momentum method is that the detailed flow processes need not be described. The effect of the perforated pipe design parameters, such as the porosity, the area ratio, the spacing between holes, and L/D ratio, is included in the two performance parameters, k and f. The two parameters can be measured by experiment. In addition, the momentum method avoids the difficulty with applying a Bernoulli equation to the branching process lying in the ambiguity which exists in identifying a relevant streamline on which to conserve energy and estimate frictional losses.

Although the flow distribution may be approximately uniform for a properly chosen M the static pressure distribution still depends on the inlet Reynolds number. There should be a value of Re_0 for a fixed M, which makes $\Delta \bar{P}_{x0}$ uniform. Exhaust equipment into the condenser stipulates that the perforated pipe should be designed for $1/6 \leqslant M < 1/4$. These conditions impose serious limitations on the capacity and utility of the perforated pipe. This result is valuable not only to the design of the perforated pipe of a new condenser, but also provides a practical operating guide in a utility-type condenser.

The friction and momentum factor still vary along the axis of the perforated pipe because of varying flow velocity. Therefore the effect of the flow velocity on f and k should be considered in further details to improve the precision of calculation. A constant friction and momentum factor are only an approximation. Wang's studies ([2]) indicated that the predicted values on the constant factors agreed qualitively with the experimental results, but they can help to clarify the basic characteristics of the flow in the perforated pipe. On the other hand, the designer may use the simple solution obtained here as a convenient mathematical tool to perform quick performance trade evaluations. A further detailed study on varying f and k can be founded in the literature ([3]).

Conclusions

There is a general characteristic parameter $M(kD/Lf)$, which depends on the pipe design and flow properties, for all perforated pipe distributors. Four specific cases can be considered, dependant on the value of characteristics parameter m: (1) when $M \leqslant 1/4$, momentum controls: the main channel static pressure will increase in the direction of the streamline; (2) when $1/6 \leqslant M < 1/4$, the momentum effect balances friction loss: the pressure will decrease to a minimum, and then increase in the direction of flow to a positive value; (3) when $0 < M < 1/6$, friction controls: the pressure will decrease to a minimum, then increase slowly, but the static pressure difference will always be negative at the whole pipe; (4) when $M = 0$, a limiting case: it denotes the limit when the ratio of the length and the diameter of friction factor is infinite. The predicted values agreed qualitively with the experimental results.

In any case it would appear that computations of the flow distribution from a constant cross-section pipe distributor would be useful to the designer who wishes to provide for nearly but not perfectly uniform flow distribution. Furthermore, it is helpful to the user who hopes to know the best operational properties and conditions for utility-type perforated pipes.

Nomenclature

D = pipe distributor diameter
f = friction factor
F = cross-sectional area of the pipe
F_c = cross-sectional area of a hole
k = momentum recovery factor
L = length of the pipe distributor
M = characteristic parameter ($=kD/Lf$)
n = number of holes
P = pressure of the pipe
$\bar{P}$ = dimensionless pressure in the pipe ($=p/\rho w_0^2$)
U_c = hole velocity
W = velocity in the pipe
W_0 = the pipe velocity at the open end
W_x = the pipe velocity at the point x
W_c = axial component of the flow velocity at the hole inlet
x = axial coordinate
$\bar{X}$ = dimensionless axial coordinate
ρ = fluid density
β = average velocity ratio in the pipe
τ_w = wall shear stress

References

[1] Sebald, J. F., Phillips, N. A., and Haman, L. L., 1982, ''Recommenced Guidelines for the Admission of High-Energy Fluids to Steam Surface Condenser,'' EPRI-CS-2251.
[2] Wang, J., 1995, ''Theory of Radial Flow Distributor and Characteristics of a Parallel-Multiple Jet,'' Ph.D. thesis, East China University of Science and Technology, Shanghai.
[3] Wang, J., Ge, X., and Wu, D., 1998, ''Progress of Flow in Manifold,'' Adv. in Mech., **28**, No. 3, pp. 392–401 (in Chinese).
[4] Bassiouny, M. K., and Martin, H., 1984, ''Flow Distribution and Pressure Drop in Plate Heat Exchangers-I: U-Type Arrangement,'' Chem. Eng. Sci., **29**, No. 4, pp. 693–700.
[5] Chang, Chengfang, Zhu, Zibin, Xu, Maosheng, and Zhu, Bingchen, 1979, ''An Investigation of Design on the Uniform Fluid Distribution for Radial Flow Reactors,'' J. Chem. Ind. Eng., No. 1, pp. 67–90 (in Chinese).
[6] Jin, Yong, Yu, Zhiqing, Su, Zhufan, and Peng, Binpu, 1984, ''Investigations on Branched Pipe Distributors for Fluidized Beds (II) Determination of Design Parameters for Branched Pipe Distributors,'' J. Chem. Ind. Eng., **35**, No. 3, pp. 204–213 (in Chinese).