Journal of Engineering for Gas Turbines and Power

Published Quarterly by ASME

VOLUME 128 • NUMBER 4 • OCTOBER 2006

TECHNICAL PAPERS

Gas Turbines: A Review Paper

Gas Turbines: Aircraft Engines

Gas Turbines: Combustion & Fuels

Gas Turbines: Controls, Diagnostics & Instrumentation

Gas Turbines: Cycle Innovations

Gas Turbines: Electric Power

Gas Turbines: Industrial & Cogeneration

(Contents continued on inside back cover)

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.

Canadian Goods & Services Tax Registration #126148048

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

(**Contents continued**)

Journal of Engineering for Gas Turbines and Power **OCTOBER 2006**

Volume 128, Number 4

Gas Turbines: Structures and Dynamics

Internal Combustion Engines

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

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

A Review of Wave Rotor Technology and Its Applications

Pezhman Akbari
e-mail: akbari@iupui.edu

Razi Nalim
e-mail: mnalim@iupui.edu

Department of Mechanical Engineering,
Purdue School of Engineering and Technology,
Indianapolis, IN 46202-5132

Norbert Mueller
Department of Mechanical Engineering,
Michigan State University,
2455 Engineering Building,
East Lansing, MI 48824-1226
e-mail: mueller@egr.msu.edu

The objective of this paper is to provide a succinct review of past and current research in developing wave rotor technology. This technology has shown unique capabilities to enhance the performance and operating characteristics of a variety of engines and machinery utilizing thermodynamic cycles. Although there have been a variety of applications in the past, this technology is not yet widely used and is barely known to engineers. Here, an attempt is made to summarize both the previously reported work in the literature and ongoing efforts around the world. The paper covers a wide range of wave rotor applications including the early attempts to use wave rotors, its successful commercialization as superchargers for car engines, research on gas turbine topping, and other developments. The review also pays close attention to more recent efforts: utilization of such devices in pressure-gain combustors, ultra-micro gas turbines, and water refrigeration systems, highlighting possible further efforts on this topic. Observations and lessons learnt from experimental studies, numerical simulations, analytical approaches, and other design and analysis tools are presented. [DOI: 10.1115/1.2204628]

Introduction

Oscillatory and pulsatile fluid motion is ubiquitous in Nature, yet is relatively poorly studied by engineers despite the invention of cyclically operating engines and machines. The potential for utilizing unsteady flows has been recognized since the early twentieth century, but has been neglected as long as substantive improvements could be made to conceptually simple semi-static devices, steady-flow devices, or crypto-steady devices (having flow that is steady in a particular frame of reference, e.g., turbomachines). Further, the inherent nonlinearity of large-amplitude wave phenomena in compressible fluids necessitates detailed flow calculations, which until recently were too laborious, expensive, or imprecise. By understanding and exploiting complex unsteady flows, significantly better engines and thermodynamic cycles can be enabled for various applications.

Shock tubes, shock tunnels, pulse combustors, pulse detonation engines, and wave rotors are a few examples of unsteady-flow devices. The basic concept underlying these devices is the transfer of energy by pressure waves. By generating compression and expansion waves in appropriate geometries, wave machines can transfer energy directly between different fluids without using mechanical components such as pistons or vaned impellers. The major benefits of these unsteady-flow machines is their potential to generate large pressure changes in short time or distance [1,2], and to tolerate transient peak fluid pressures and temperatures that exceed continuous exposure limits. Furthermore, wave compression is a relatively efficient process at moderate pressure ratios as shown in Fig. 1, where shock isentropic efficiency η_{Shock} (red) is compared with compressor isentropic efficiency $\eta_{\text{Compressor}}$ (green) and subsonic diffuser isentropic efficiency η_{Diffuser} (blue). Figure 1 shows variations of these parameters as functions of the pressure gain p_2/p_1 obtained by a moving shock wave in a frictionless channel, by a compressor with different values of polytropic efficiencies, and by a diffuser with different values of total pressure drop expressed by p_{t2}/p_{t1}, respectively. The comparison reveals that for the same pressure gain p_2/p_1, the ideal shock compression efficiency may significantly exceed the efficiency obtained by a typical diffuser or compressor. Flow friction effects would lower the efficiency of wave devices [3] and reduces their efficiency advantage (not shown in Fig. 1), but the relative advantage is expected to persist. As an example, Ref. [4] has investigated a feasibility study of replacing conventional diffusers used in centrifugal compressors with the wave augmented diffuser.

Wave Rotor Machines

The essential feature of a wave rotor is an array of channels arranged around the axis of a cylindrical drum. As schematically shown in Fig. 2, the drum rotates between two stationary end plates, each of which has a few ports or manifolds, controlling the fluid flow through the channels. Through rotation, the channel ends are periodically exposed to differing port pressures, initiating compression, and expansion waves within the wave rotor channels. The number of ports and their positions vary for different applications. By carefully selecting their locations and widths to generate and utilize wave processes, a significant and efficient transfer of energy can be obtained between flows in the connected ducts. Thus, pressure is exchanged dynamically between fluids by utilizing unsteady pressure waves. Unlike a steady-flow turbomachine that either compresses or expands the fluid, the wave rotor accomplishes both compression and expansion within a single component. To minimize leakage, the gap between the end plates and the rotor has to be very small, but without contact under all operating and thermal expansion conditions. An inverted design with stationary channels and rotating ports is also possible [5]. Such a configuration may be preferred for laboratory investigations for easy flow measurement in the channels where the important dynamic interactions take place. In practical design, this arrangement is mechanically inconvenient [6].

Most designs use straight axial channels, but curved channels can be used to create a "wave turbine" that produces shaft power. With axial channels and matched port flow alignment, the power required to keep the rotor at a correctly designed speed is negligible [6,7], as it only needs to overcome rotor windage and friction. In such a configuration, the rotor may be gear or belt driven or preferably direct driven by an electrical motor (not shown). Alternatively, a self-driving configuration, known as the "free-running rotor," can drive itself by using port flow incidence on channel walls to turn the rotor [8,9].

There are several important advantages of wave rotor machines relative to competing turbomachines particularly for straight-channel rotors with no shaft power transfer. Wave rotor flows can respond on the time scale of pressure waves with no rotational inertia lag. Their rotational speed is low compared with turboma-

---

Submitted to ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received December 7, 2004; final manuscript received January 19, 2006. Review conducted by R. P. Shreeve.

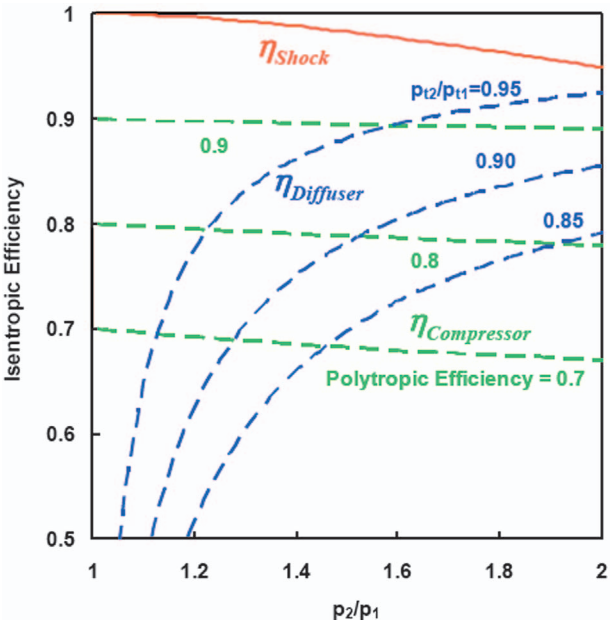


Fig. 1 Shock wave, compressor, and diffuser isentropic efficiencies as functions of pressure gain

chines, unless designed for shaft power generation, which should result in low material stresses. However, the tip shroud offsets this advantage somewhat, and the design must safeguard against fatigue-induced failures in surfaces subjected to cyclic pressure fluctuations. Wave rotor geometry can be mechanically simpler than those of turbomachines, allowing inexpensive manufacture. In addition the rotor channels are less prone to erosion damage than the blades of turbomachines. This is mainly due to the lower velocity of the working fluid in the channels, which is about one-third of what is typical within turbomachines [6] and the absence of flow turning. Another important advantage of wave rotors is their self-cooling capability. In heat engine applications, the rotor channels pass both cool air (being compressed) and hot gas (being expanded) in the cycle at least once per rotor revolution, alternating faster than thermal diffusion rates, allowing peak cycle temperature above materials limits. The rotor temperature equilibrates between the temperature of the cooler air and the hotter gas, but may retain axial and radial temperature variation that limits strength and distorts sealing surfaces.

Despite generally attractive features, several challenges have impeded the extensive commercial appearance of wave rotors. Numerous research efforts have been carried out during the past century to understand the complex unsteady flow and creatively select the best wave rotor configuration for a particular application. The obstacles have been mainly of a mechanical nature, such as sealing and thermal expansion issues, as mentioned frequently in this review. Nevertheless, continued impetus for energy efficiency, diminishing advances of older technology, and marketplace changes have stimulated new interest in wave rotor technology.

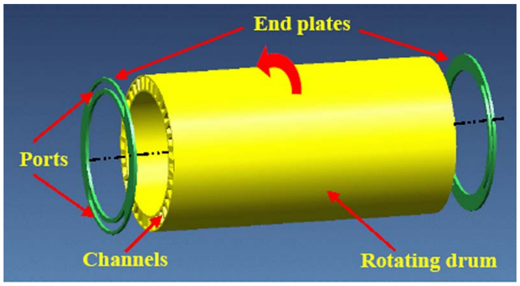


Fig. 2 Schematic configuration of a typical wave rotor

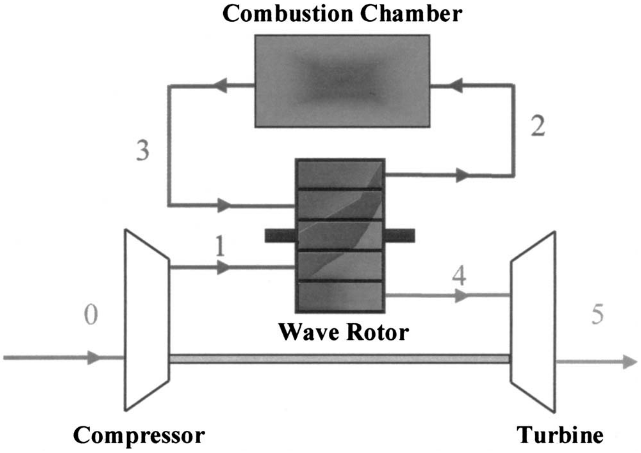


Fig. 3 Schematic of a gas turbine topped by a through-flow four-port wave rotor

Four-Port Pressure-Exchange Wave Rotor Examples. A variety of wave rotor configurations have been developed for different applications. The number and azimuthal location of the wave rotor ports along with heat addition schemes distinguish them for different purposes. As will be shown in the next section, four-port configurations have been used mainly as superchargers for internal combustion engines. Three-port wave rotors have been employed in pressure dividers and pressure equalizers in which the pressures of different fluids are increased or reduced. Two-port, four-port, five-port, and nine-port wave rotors have been investigated for gas turbine engine topping applications, including some with on-board combustion. As an application of current interest, a four-port pressure exchange wave rotor integrated into a gas turbine cycle is briefly discussed below to illustrate wave rotor operation and options.

Figure 3 shows a schematic of a gas turbine cycle using a four-port wave rotor. Following the flow path shown in Fig. 3, air from the compressor enters the wave rotor (state 1) and is further compressed inside the wave rotor channels. After the additional compression of the air in the wave rotor, it discharges into the combustion chamber (state 2). The hot gas leaving the combustion chamber (state 3) enters the wave rotor and compresses the air received from the compressor (state 1). To provide the energy transfer to compress the air, the burned gas partially expands in the wave rotor en route to the turbine (state 4). In this configuration, combustion takes place at a higher pressure and temperature than in a conventional gas turbine engine with the same compressor exit state, while being limited to the same turbine inlet temperature. The turbine inlet total pressure is typically 15% to 20% higher than pressure of the air delivered by the compressor [10]. This pressure gain is in contrast to the untopped engine where the turbine inlet pressure is always lower than the compressor discharge pressure, due to the pressure loss across the combustion chamber. As a result of the wave rotor pressure gain, more work can be extracted from the turbine, increasing overall engine thermal efficiency and specific work. For instance, a study by Rolls-Royce Allison has predicted [11,12] significant performance improvements for both design point and off design operating conditions of the Allison model 250 turboshaft engine topped by a four-port wave rotor, as shown in Fig. 4. It compares specific shaft horsepower and decreases in specific fuel consumption for the topped and baseline engines as a percent improvement for the off design points.

In the above-described wave rotor, both gas and air inlet ports

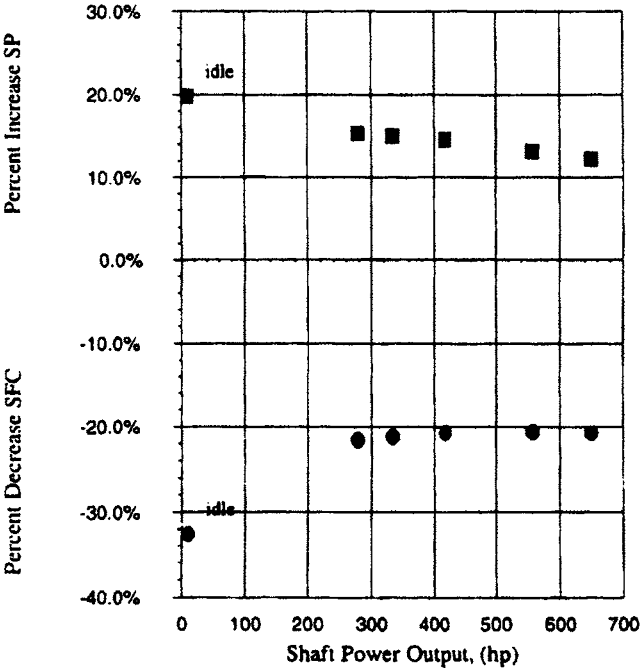


Fig. 4 Comparison of off-design performance to baseline engine performance, taken from Ref. [11]

are located on one side of the rotor while and the outlet ports are located on the other side of the rotor. This configuration is known as the through-flow (TF) wave rotor in the literature. Alternatively, another type of wave rotor has been designed where the fresh air enters and exits at the same end of the rotor (air casing) while the burned gas enters and exits the rotor at the other end (gas casing). This configuration is called the reverse-flow (RF) wave rotor, shown in Fig. 5. These two configurations may provide identical topping and overall performance enhancement, but they differ substantially in their internal processes. In the TF four-port wave rotor, both hot gas and relatively cold air traverse the full length of the rotor, keeping the wall at a relatively uniform intermediate temperature. This self-cooling feature of TF wave rotors has prompted interest in them for gas turbine engine topping applications where gas temperatures are high. The RF configuration does not inherently result in such a self-cooled rotor. The cold air never reaches the other end of the rotor, as seen from Fig. 5. As a result, the air side of the rotor is relatively cool while the gas side of the rotor is relatively hot. Thus, the RF configurations have been mostly used in the relatively low-temperature application of engine supercharging although such a configuration for gas turbines has been also investigated [11–14]. The General Electric Company has obtained experimental data on a gas turbine engine enhanced by a RF wave rotor [15].

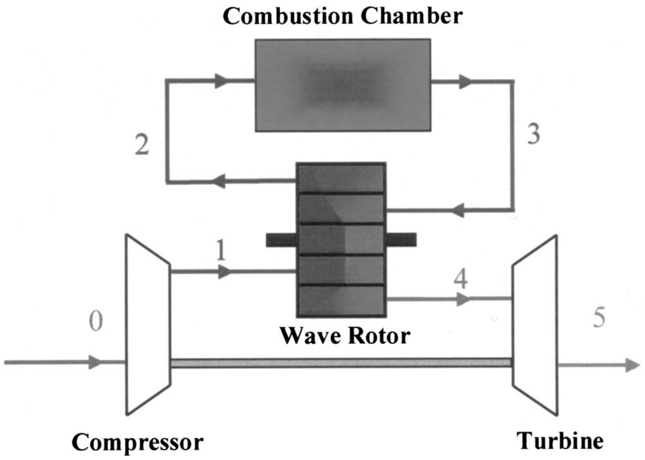


Fig. 5 Schematic of a gas turbine topped by a reverse-flow four-port wave rotor

Inner Workings of a Pressure-Exchanger. Two basic fluid-exchange processes usually happen at least once per revolution: the high-pressure (charging) process and the low-pressure (scavenging) process. In the high-pressure process, compression waves transfer the energy directly from a fluid at a higher pressure (driver fluid) to another fluid at a lower pressure (driven fluid). In the low-pressure process, the driver fluid is scavenged from the rotor channels, generating expansion waves that allow ingestion of a fresh low-pressure fluid into the rotor channels.

The wave process occurring inside the wave rotor channels is customarily illustrated by the wave diagram (space-time diagram), where the circular motion of the rotor channels is represented on paper in developed view by a straight translatory motion. It describes the rotor internal operation by tracing the trajectories of the waves and gas interfaces. The wave diagram is very useful for visualizing the wave process occurring inside the channels and also for explaining wave rotor design parameters, i.e., port opening and closing times and their locations. The utility of the wave diagram is analogous to that of a velocity diagram for a conventional turbine or compressor.

Figure 6 taken from a NASA publication [16] presents wave diagrams for the TF (left) and the RF (right) four-port wave rotors, showing the sequence of events occurring during one cycle within the channels moving in the upward direction. The journey of a channel of the wave rotor is periodic. The top of each wave diagram is therefore considered to be looped around and joined to the bottom of the diagram. This presents a fundamental requirement in the simulation and design of wave rotors.

The events occurring in one cycle of a TF four-port wave rotor are now described. As shown in Fig. 6, the process begins in the bottom part of the left wave diagram, where the channel is closed at both ends and contains medium-pressure gas at state V. As the channel gradually opens to the relatively low-pressure outlet port, an expansion fan originates from the leading edge of the outlet port and propagates into the channel, discharging the gas to the turbine. The expansion fan reflects off the left wall and reduces the pressure and temperature inside the channel further. The inlet port at left opens shortly (time t_c), and this depression draws fresh air provided by the compressor into the channel. When the reflected expansion fan reaches the outlet port (t_a), it slows the outflow and reflects back as compression waves, while the outlet port then closes (t_b) and halts the flow inside the channel. The compression waves form a single shock wave as they travel toward the inlet port. As the shock wave reaches the upper corner of the inlet port (t_d), it closes gradually trapping both residual gas (D) and fresh air (C) at state Q.

The above sequence of events is called the low-pressure part of the cycle (scavenging process). Its purpose is to discharge a relatively high-pressure gas into the turbine, partially purge the rotor channels, and ingest fresh low-pressure air received from the compressor. In the high-pressure part of the cycle (charging process) that follows, the rotor channels are first exposed to high-pressure burned gas from the combustion chamber. This hot gas (driver) penetrates the channel triggering a shock wave from the lower corner of the inlet port (t_e). The shock wave runs through the channel and causes an abrupt rise of pressure inside the channel. As the shock wave reaches the right end of the channel, the outlet port opens (t_g) and its lower edge originates a reflected shock wave that propagates back into the channel. The twice-compressed flow comprising both air and once-burned gas behind the reflected shock wave leaves the wave rotor toward the combustion chamber with total pressure sufficient to overcome com-

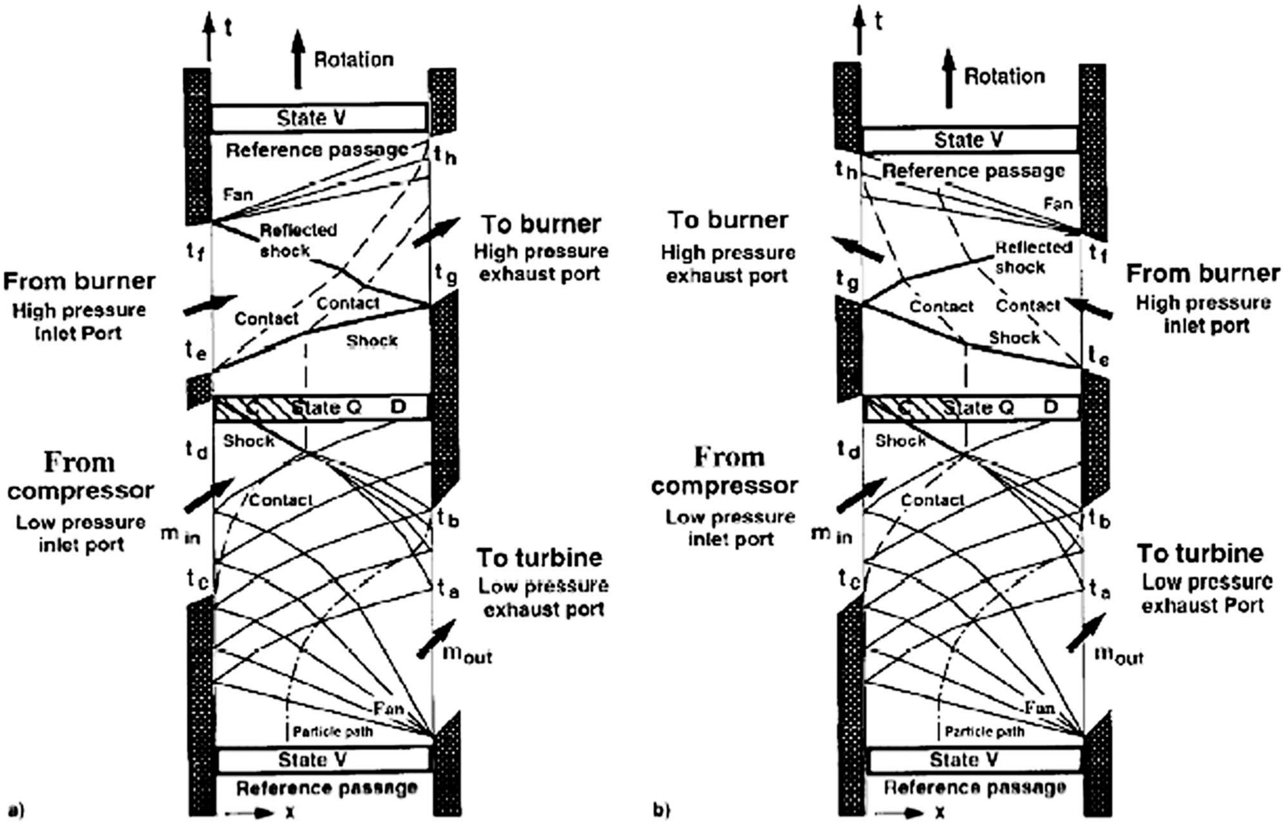


Fig. 6 Wave diagrams for through-flow (left) and reverse-flow (right) four-port wave rotors, taken from Ref. [16]

bustor pressure loss. Detailed fluid flow investigations have suggested that approximately 30% to 50% of burned gas is recirculated to the combustion chamber in the TF configuration [17]. A favorite case is considered when the closure of the gas inlet port is timed (t_f) with the arrival of the reflected shock wave. At this moment, an expansion fan originates from the upper corner of the inlet port and brings the channel flow to rest. When the expansion fan reaches the end of the channel, the outlet port closes (t_h) and the flow in the rotor channels stops with the same state V as at the starting state of the cycle. It is now ready to be discharged into the turbine by the low-pressure process, explained above.

The RF configuration will not be presented in detail, but the right-hand wave diagram of Fig. 6 illustrates most features. The cycle begins with the channel containing both hot gas and a buffer layer separated by a contact surface. The low-pressure scavenging part of the cycle is similar to the TF configuration. The high-pressure part has the driver gas inlet on same side as the turbine outlet, allowing only fresh air to be sent to the burner. The buffer gas oscillates back and forth in the channel, never leaving it except by gradual mixing at the contact surfaces. Buffer gas or residual gas can be avoided in pressure-exchangers that do not seek a net pressure gain between the low-pressure ports (such as a supercharger) or have a fifth lower pressure discharge port.

Historical Review

The following roughly chronological review attempts to summarize the developments in methods and designs of major participating institutions that began before 1985. Many applications were considered, and two or three emerged as commercial or unique solutions, even with simplified methods of predicting complex unsteady flow, perhaps because efficiency was not critical in those applications.

The Early Work (1906–1940). The earliest pressure exchanger proposed by Knauff in 1906 [18] did not employ the action of pressure waves. It consisted of a cellular drum that rotates between two end plates containing several ports through which flows with different pressures enter and leave, exchanging their pressure. Knauff initially described curved rotor blades and inclined stator nozzles to provide output shaft power (pressure exchange engine). Reported by Pearson [19], Knauff in his second 1906 patent [20] and Burghard in 1913 [21] proposed a simpler device in which pressure exchange takes place in long narrow channel configurations (pressure exchanger) known later as the Lebre machine following Lebre's patent in 1928 [22]. Around 1928, Burghard proposed the utilization of pressure waves [23] in what was termed the "dynamic pressure exchanger" to distinguish it from the previous "static pressure exchangers." Here, the term "dynamic" implies the utilization of pressure waves in both compression and expansion processes taking place inside the rotor channels; hence our preferred term "wave rotor." Limited understanding of unsteady fluid mechanics delayed development of the wave rotor concept [24] until World War II.

The Brown Boveri Comprex® Pressure Wave Supercharger (1940–1989). Brown Boveri Company (BBC), later Asea Brown Boveri (ABB), in Switzerland has a long history in wave rotor technology. As reported by Meyer [25], BBC in the early 1940s designed a wave rotor as a topping stage for a 1640 kW (2200 hp) British Railways locomotive gas turbine [26–29]. An 80% power boost and a 25% efficiency increase was expected, based on the patents of Seippel [30–33]. The wave rotor had 30 channels rotating at 6000 rpm, with two opening ports on each side through which air and gas entered and left. It had originally shown a pressure ratio up to 3:1 and total efficiency of 69% in tests during 1941–1943, which correspond to 83% efficiency for each compression and expansion process [25]. The first wave rotor worked

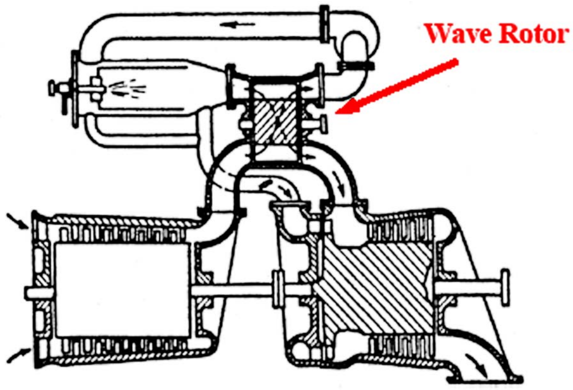


Fig. 7 Wave rotor as a topping stage for the locomotive gas turbine, taken from Ref. [25]

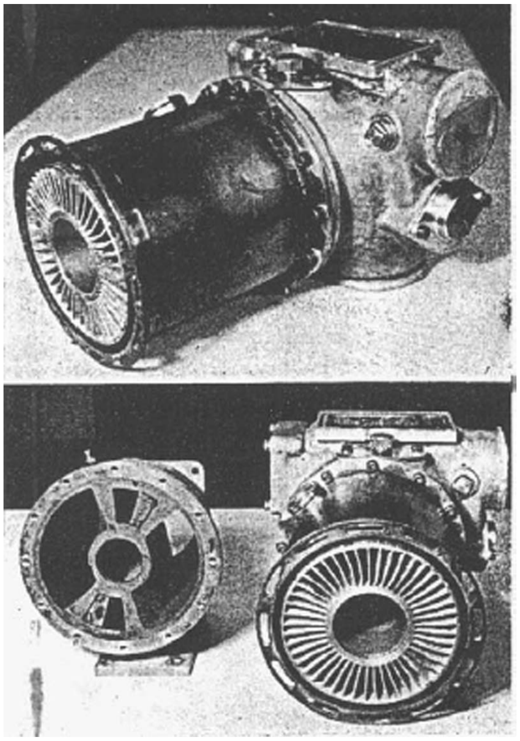


Fig. 8 The Comprex®, taken from Ref. [42]

satisfactorily, proving the concept of wave rotor machines. However, its performance, when installed in the engine as shown in Fig. 7, was far from expectations, apparently due to mismatched design and crude integration [28].

Seippel's work, and growing practical knowledge led to the notion of using the wave rotor as a pressure wave supercharger for diesel engines, first by the ITE Circuit Breaker Company in the U.S. [34–36]. In an effort jointly sponsored by the U.S. Bureau of Aeronautics and ITE, supervised by Kantrowitz of Cornell University and Berchtold of ITE, the first units were manufactured and tested on vehicle diesel engines between 1947 and 1955. As a result of this success, a cooperative program with BBC was started in 1955. As a manufacturer of superchargers, BBC pursued the development of pressure-wave superchargers for diesel engines [37], collaborating with the Swiss Federal Institute of Technology (ETH Zurich). While the first prototype was installed in a truck engine in 1971 [38], the supercharging of passenger car diesel engines was started in 1978 [39,40] with a first successful test on an Opel 2.1 l diesel engine [40,41]. Given the trade name Comprex®, (Fig. 8), its port arrangement indicates the use of two operating cycles per revolution, shortening the rotor length and reducing thermal loads. The main advantage of the Comprex® compared with a conventional turbocharger is its rapid response to driver demand. Light weight and compact size make this device attractive for supercharging small engines (below about 75 kW or 100 hp) [42,43]. By 1987, the first wide application of the Comprex® in passenger cars appeared in the Mazda 626 Capella [8,44], and ultimately in 150,000 Mazda diesel passenger cars [45]. TheComprex® has been also tested successfully on vehicles such as Mercedes-Benz [9], Peugeot and Ferrari [37], and on heavy-duty engines.

The successful development of the Comprex® has been enabled by efforts of other researchers including: Gyarmathy [7], Burri [46], Wunsch [47], Croes [48], Summerauer [49], Kollbrunner [50], Jenny [51], Keller [52], Rebling [53], and Schneider [54], others at BBC [55–68] and elsewhere [69–79]. By the end of the 1980s, the Comprex® activity was transferred to Mazda in Japan [24,80], when researchers at ABB returned to the idea of utilizing wave rotor technology for gas turbine applications [81,82]. Nalim has reviewed ABB's recent efforts in developing both pressure-exchange and combustion wave rotors in Ref. [83,84].

During 1990s, a few groups continued research on pressure wave superchargers. Nour Eldin and associates at the University of Wuppertal in Germany developed a numerical method using the theory of characteristics [85–91]. Piechna et al. at the Warsaw University of Technology developed one-dimensional and two-dimensional numerical codes [92–98]. Oguri et al. at Sophia University in Japan performed measurements on a gasoline engine supercharged by the pressure wave supercharger [99]. Guzzella et al. [43,100–104] at ETH in Switzerland developed a control-oriented model that describes the engine supercharged by pressure wave devices, with special emphasis on the modeling of transient exhaust gas recirculation. The experimentally validated model has introduced an optimized strategy to operate a supercharged engine with good drivability. Finally, an investigation of Comprex® supercharging on diesel NO_x emissions has proved fruitful in Turkey [105].

To date, the Comprex® has been the most commercialized of the wave rotor devices. The Comprex® development by BBC/ABB also has established fabrication techniques for wave rotors in commercial quantities and produced a mature and reliable machine for internal combustion engine supercharging. For this application, BBC/ABB has solved difficult development challenges in sealing against leakages, noise, and thermal stress. For instance, leakage was kept to an acceptable level by enclosing the rotor in a pressurized casing and using a rotor material with a low thermal expansion coefficient over the operating temperature range [37]. Furthermore, off-design performance over the engine speed range was improved by using pockets in the end plates to control wave reflections [67].

In recent years, Swissauto WENKO AG in Switzerland has developed a more sophisticated version of the pressure wave supercharger [45], known as the Hyprex®, for small gasoline engines. It benefits from the new ETH control features, enabling higher pressure ratios at low engine speeds, further reduced noise levels, and improvement of compression efficiency at medium or high engine speeds. The Hyprex® has been successfully demonstrated with a two-cylinder gasoline engine in a modified Renault Twingo, achieving very low specific fuel consumption and low emissions.

Cornell Aeronautical Laboratory and Cornell University (1948–2001). Inspired by the cooperation with BBC in the late 1940s, work on unsteady-flow concepts was initiated at Cornell Aeronautical Laboratory (CAL). Among several novel concepts including development of energy exchangers for gas turbine cycles and various stationary power applications [106], the CAL

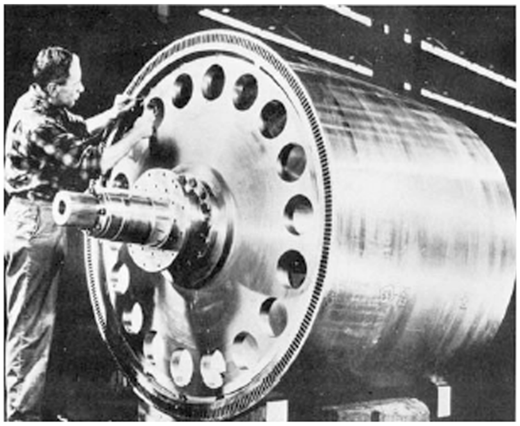


Fig. 9 Photograph of the CAL Wave Supercharger, taken from Ref. [6]

Wave Superheater was built in 1958 and utilized until 1969 [29]. The 2 m diameter Wave Superheater, photographed in Fig. 9, used heated helium as the low molecular weight driver gas to provide a steady stream of high-enthalpy air for a hypersonic wind tunnel test facility. It compressed and heated air to more than 4000 K and up to 120 atm for run times as long as 15 s. The CAL Wave Superheater was a landmark demonstration of the high temperature capabilities of wave rotor devices [29,106].

Around 1985, Resler, a former member of the CAL Wave Superheater team, recommenced wave rotor research at Cornell University. His efforts with his group led to the development of new wave rotor concepts and analytic methods for three-port wave rotor diffusers [107], double-wave rotor cycles [108], five-port wave rotors [108–113], and supersonic combustor aircraft engines using wave rotors [114]. Numerical modeling indicated significant potential for reducing NO_x in gas turbine engine applications by using a wave rotor for rapid expansion of fuel-rich combustion products. Figure 10 illustrates a double wave rotor in a gas turbine cycle. The idea of using a compound unit consisting of two (or multiple) wave rotors, one supercharging the other, is also reported in an early German patent by Müller in 1954 [115], as stated by Azoury [42].

Power Jets Ltd (1949–1967). In parallel with, but independent of Seippel's efforts in 1940s, Jendrassik, former chief engineer of the Gantz Diesel Engine Company of Budapest, was working on the development of wave rotor machines for gas turbine applications [28,116–118]. He quickly realizes its benefit for aircraft engines, proposing the wave rotor as a high pressure topping stage for early aircraft engines [119,120]. His ideas stimulated the British government-controlled Power Jets, Ltd. of jet-engine fame to become active in the wave rotor field in 1949. Initial interest in IC engine supercharging later extended to several other applications including air-cycle refrigerators, gas turbines, pressure equalizers, and dividers [6,24,28,118]. For instance, two prototype air-cycle refrigerators using wave rotors were commissioned in gold mines in India and South Africa, with the advantage of low weight and bulk. After Jendrassik's death in 1954, theoretical and experimental work continued at Imperial College, University of London, directed by Spalding and Barnes, and at Ricardo [28,121]. The experimental divider test rig at Imperial College is shown in Fig. 11. Power Jets, Ltd. efforts are detailed in company reports listed in Ref. [6].

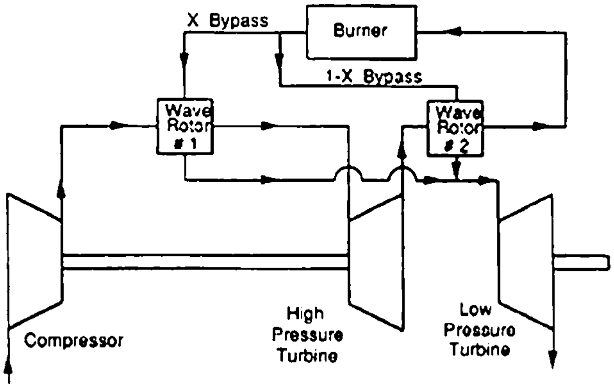


Fig. 10 Schematic of a double wave rotor cycle, taken from Ref. [108]

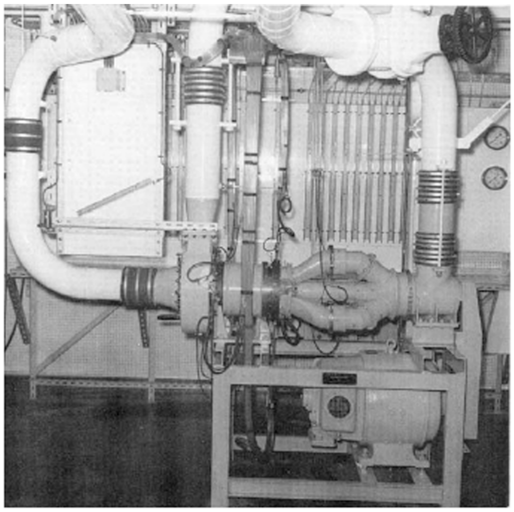


Fig. 11 The experimental divider test rig at Imperial College, taken from Ref. [6]

Spalding of Imperial College pioneered computational methods for wave rotors considering the effects of heat transfer and friction. It utilized novel features to ensure solutions free from instabilities and physical improbabilities [28]. At the time, manual analysis using characteristics was quite tedious, and could not easily account for loss effects. Based on his numerical model, a computer program was developed by Jonsson [122], and it was successfully applied to pressure exchangers [123–125]. Spalding's students, Azoury [126] and Kentfield [127], continued their efforts on different theoretical aspects of pressure exchangers [6,24,28,42,118,125,128–131] despite the dissolution of Power Jets, Ltd. in 1967 [28].

Ruston-Hornsby Turbine Company: The Pearson Rotor (Mid 1950s–1960). Also in the UK of the mid-1950s, the Ruston-Hornsby Turbine Company, manufacturer of diesel engines and industrial gas turbines, supported the construction and testing of a different kind of wave rotor designed by Pearson [132,133]. This unique wave rotor, known as the wave turbine engine or simply the wave engine, has helical channels that change the direction of the gas flows producing shaft work similar to a conventional turbine blade. Pearson designed and tested his wave rotor, shown in Fig. 12, in less than a year. The rotor has a 23 cm (9 in.) diameter and a 7.6 cm (3 in.) length. The engine apparently worked successfully for several hundred hours in a wide range of operating conditions (e.g., 3000–18,000 rpm) without variable porting, and produced up to 26 kW (35 hp) at its design point with a cycle peak temperature of 1070 K and a thermal efficiency of around 10%. Better performance seemed possible with improvements to overcome leakage and incomplete scavenging. The design of the engine was based on complex wave diagrams using the method of characteristics that accounted for internal wave reflections, and extra ports and injection nozzles to control and cancel unwanted

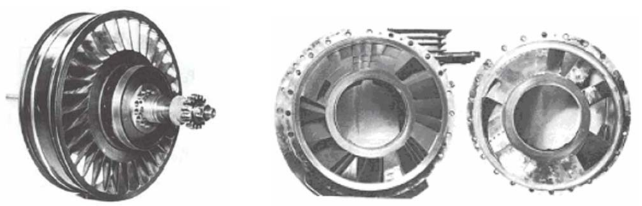


Fig. 12 The Pearson rotor (left) and rear and front stator plates (right), taken from Ref. [132]

reflected waves. The engine had a length of only one-third of its diameter despite having only one cycle per revolution [19]. The sealing and bearings were carefully adapted, considering rotor thermal expansion. Eventually, the engine was wrecked due to over-speeding from an improperly connected fuel line, and the project was tragically canceled when success seemed so close. Despite the technical success achieved, Pearson failed to attract additional funding for this radical idea.

In the early history of wave rotor technology, the Pearson rotor and the Comprex® have performed efficiently over a wide range of operating conditions [28,37,117], demonstrating good off-design performance, while the Wave Superheater was an equal success at its narrow purpose. Yet, the Pearson rotor is notable for producing a significant power output in addition to being a successful pressure exchanger.

General Electric Company (1956–1963). Contemporaneously, General Electric Company (GE) in the U.S. initiated a wave rotor program in 1956 [15]. The work was motivated by earlier work at NASA Langley initiated by Kantrowitz and continued by Huber [134] during the development of a wave engine in the early 1950s and later in 1954–1956 developing pressure gain combustors [15]. GE studied this new configuration of wave rotor in which combustion took place inside the rotor channels (internal combustion wave rotors). Such an arrangement eliminates the external combustion chamber used in the gas turbine cycle, promising significantly lower weight, less ducting, and a compact size. In the period of 1956 to 1959, the methods used at NASA were analyzed, improved and applied to the design and fabrication of the first internal combustion wave rotor. As reported by Weber [2], the test rig was first tested at the California Advanced Propulsion Systems Operation (CAPSO) of GE. After 20 s of operation, the rotor seized between the end plates. The test demonstrated the difficulty of clearance control between the end plates and rotor during uneven thermal expansion. While the running clearance between the end plates and rotor must be kept as small as possible, the rotor tends to expand thermally due to hot gases in the rotor, while end plates may warp. This is an especially challenging problem in the design of combustion wave rotors with localized heating and temperature gradients. Henceforth, GE resorted to inferior rubbing seals, and tested only pressure-exchange configurations from 1960 to 1961 [15]. Despite flow leakage, respectable wave rotor overall pressure ratios of 1.2 to 1.3 corresponded to overall temperature ratios of 1.9 to 2.6 were achieved, measured between low-pressure inlet and outlet ports. Meanwhile, a feasibility study was initiated for substituting compressor stages of a T-58 GE-06 engine with a wave rotor. It showed a considerable reduction in overall engine weight and cost, and a 15% reduction in specific fuel consumption rate, motivating a conceptual design layout of such an advanced engine.

GE also pursued a wave turbine engine. Over the period from 1961 to 1963, Klapproth and his associates at GE in Ohio fabricated and tested a wave engine using air-gap seals. An ideal wave diagram of this engine is shown in Fig. 13. The engine worked continuously, but it did not produce the anticipated net output power. It is believed that insufficient attention was given to account for internal wave reflections, thus, the flow field calculations were inaccurate [37]. Simplifications were unavoidable at that time and generation of wave diagrams by hand required considerable time and effort and small design changes necessitated lengthy recalculations. Although the Klapproth rotor did not produce the expected performance, it clearly demonstrated the possibility of the complete energy conversion within the wave rotor. GE development of the wave rotor was canceled in 1963, with speculated reasons including a shift toward space exploration and rocket propulsion [2], and GE's commitment to pursue large engine development exclusively [15,134].

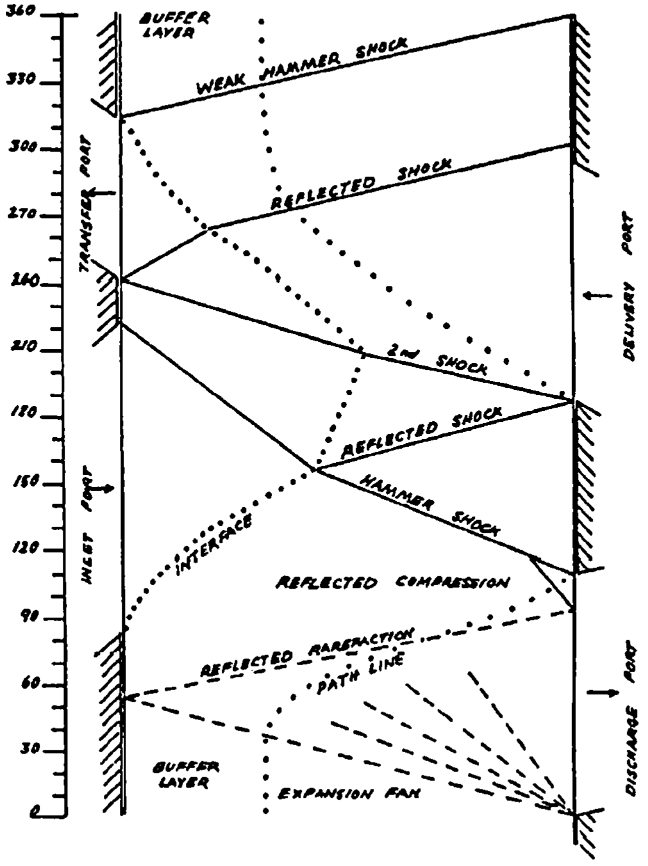


Fig. 13 Ideal wave diagram of the Klapproth rotor, taken from Ref. [15]

General Power Corporation (Mid 1960s–1984). In the mid 1960s, General Power Corporation (GPC) started a wave rotor program originally intended for a vehicle engine application [37]. Over a period of about 20 years, GPC spent considerable time and money to successfully design and develop wave rotors. The work was initially supported by Ford Motor Company and later by the Department of Energy (DOE) and the Defense Advanced Research Projects Agency (DARPA). Unfortunately, the GPC work is poorly documented [135]. As stated by Taussig [37,117], while the GPC rotor shared some of the features of the Klapproth and Pearson rotors, it differed in several aspects. Intended to produce reactive shaft power utilizing curved blades, its performance suffered from excessive blade curvatures, lack of control of reflected waves within the device, and the absence of any strong impulsive loading of the rotor from inlet manifolds to produce shaft work. The latter was in contrast with the Pearson rotor that relied heavily on impulsive loading of the rotor blades to achieve power output. Furthermore, the GPC rotor had inadequate off-design performance. Although GPC developed a computer code to obviate manual wave pattern design, accurate calculations were still tedious. Ultimately, Ford withdrew its support from the wave rotor research [136] and GPC discontinued development of the wave

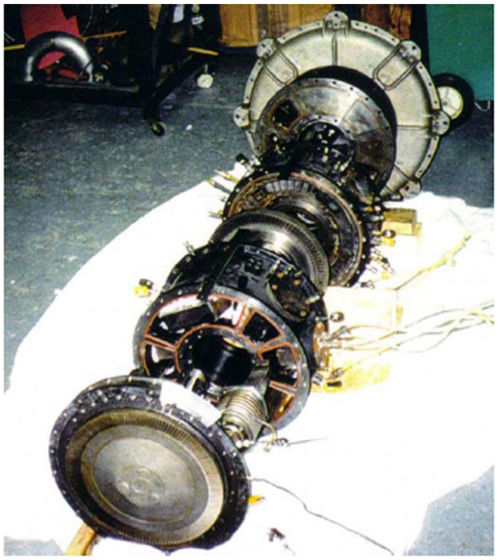


Fig. 14 Disassembled view of the GPC rotor, taken from Ref. [137]

engine in the early 1980s. In 1994, the GPC engine was transferred to the University of Florida for further tests, accurate measurements, and improvements of the seals. Figure 14 shows the disassembled view of the engine. After testing under various operating conditions, major modifications in the starting system, seals and cooling, and rotor design were identified before the engine become a well-characterized experimental testbed [137].

Rolls-Royce (1965–1972). In the mid-1960s, Rolls-Royce (RR) in the UK began numerical and experimental wave rotor research [37]. BBC cooperated with RR in the development of pressure-exchange wave rotors as topping spools in gas turbine applications [13], with Berchtold of the ETH and Spalding of Imperial College serving as consultants [24]. Considerable efforts were made to design a wave rotor as a topping stage for a small helicopter engine (Allison Model 250) [138]. The BBC-RR engine utilized a reverse-flow wave rotor incorporated into a single turbine engine. Figure 15 is a rare illustration of the RR reverse-flow wave rotor, taken from a 1965 poster. This was somewhat different from the cycle suggested by Berchtold and Lutz [71] in BBC gas-turbine-topping investigations, which employed a through-flow wave rotor integrated with both low-pressure and high-pressure turbines. BBC's interests in wave rotors at that time were mostly related to development of small gas turbines for passenger cars, beset by poor efficiencies at sizes of 100 kW and smaller [29]. While the enhanced engine operated nearly as predicated by protracted manual design methods, performance suffered from leakage [37]. Other difficulties related to the start-up, bearing durability, fuel system complications, and control are also reported [10]. The program was abruptly canceled in 1972 amid severe company financial difficulties [13]. As stated by Kentfield [24], contemporaneous rapid progress in turbomachinery technology may have disfavored high-risk projects, both at RR and GE. It is emphasized that published literature is relatively meager on efforts at large corporations like GE, ABB, and RR, and underlying business strategy is rarely revealed. Internal company records, if they survive, may contain more technical details.

Mathematical Science Northwest Inc. (1978–1985). In the late 1970s, Mathematical Science Northwest Inc. (MSNW, later Spectra Technology Inc.) investigated various applications of wave rotors [29]. Under the sponsorship of DOE and DARPA, they considered wave rotors for a broad range of stationary power systems such as magnetohydrodynamic cycles [37], combined cycles integrated with gasification plants [139], pressurized fluidized bed (PFB) power systems [140], and also propulsion and transportation applications [117]. Significant numerical and experimental efforts included developing a laboratory wave rotor [141–144], shown in Fig. 16. With diameter of 45 cm, it consists of 100 channels each with a 40 cm length. It is a four-port wave rotor with two additional small ports provided to cancel pressure waves at critical rotor locations providing more uniform port flows and a higher transfer efficiency [144]. Besides successful tests using several configurations (clearance variations, port sizes, etc.) and various operating conditions, experiments were designed to verify the scaling laws for predicting the performance of larger machines [139]. The MSNW wave rotor was initially designed based on the method of characteristics, but a one-dimensional unsteady computer code (the FLOW code) was later used for optimization [117]. Modifications led to very good agreement between the numerical and experimental results in a wide range of operating conditions. The FLOW code, which was developed specifically for both pure pressure exchanger wave rotor and wave engine analyses, used the flux-corrected transport algorithm solving Euler equations accounting for heat transfer, viscosity, gradual port opening, and flow leakage. The sensitivity of wave rotor performance to tip speed, port placement and size, inlet and outlet flow conditions, channel geometry, number of channels, leakage, and heat transfer was analyzed for both on-design and off-design

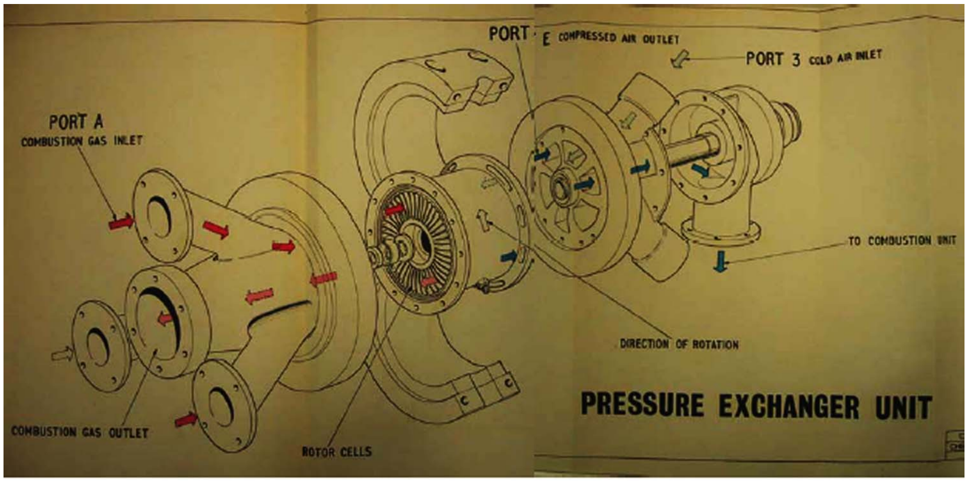


Fig. 15 Reverse-flow wave rotor of Rolls-Royce

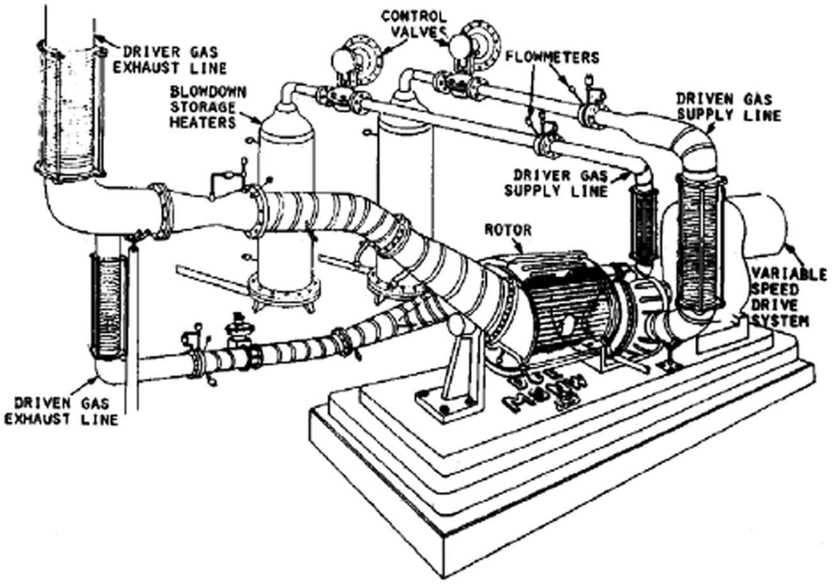


Fig. 16 Schematic of the MSNW wave rotor experimental setup, taken from Ref. [29]

conditions. It was concluded that heat transfer losses were negligible, and leakage was recognized as a key problem for efficient wave rotor operation. Numerical work was reported for a nine-port wave rotor concept to resolve the problem of nonuniform port flows and poor scavenging.

MSNW also produced preliminary wave rotor designs for a small turbofan engine generating 600 lb thrust at sea level condition [117,145], illustrated in Fig. 17. Performance calculations for both on-design and off-design flight conditions using a cycle performance code and the FLOW code simulation predicted significant performance improvements of such an enhanced engine. No new material development for such combined engines was required. The wave rotor activity at MSNW was discontinued in the mid 1980s, for reasons not reported.

Naval Postgraduate School (1981–1986). In 1981, the Office of Naval Research (ONR) agreed to monitor a joint DARPA/ONR program to evaluate the wave rotor concept and its potential application in propulsion systems [134]. Following this decision, the Turbopropulsion Laboratory (TPL) at Naval Postgraduate School (NPS), directed by Shreeve, started an extensive numerical and analytical wave rotor program. To support the accuracy of the computational results, the wave rotor apparatus formerly used by Klapproth at GE was transferred to TPL and some preliminary tests were carried out. It is reported that the rotor produced some shaft work running at approximately 5000 to 6000 rpm [146]. No further experimental details are reported.

For numerical simulations, two different approaches to the solution of the unsteady Euler equations were examined in the overall program. First, Eidelman developed a two-dimensional code based on the Godunov method to analyze the flow in wave rotor channels [147–149]. Unlike contemporary one-dimensional approaches [150], the two-dimensional code showed the effect of gradual opening of the channels. The main conclusion of these studies is that if the channels are straight, the flow remains nearly one-dimensional, which in turn leads to minimal mixing losses caused by rotational flow in the channels [151]. However, when the channel of the wave rotor is curved, even an instantaneous opening of the channel does not lead to the development of a one-dimensional flow pattern with small losses. For faster computations, a one-dimensional, first-order time-accurate code was in-

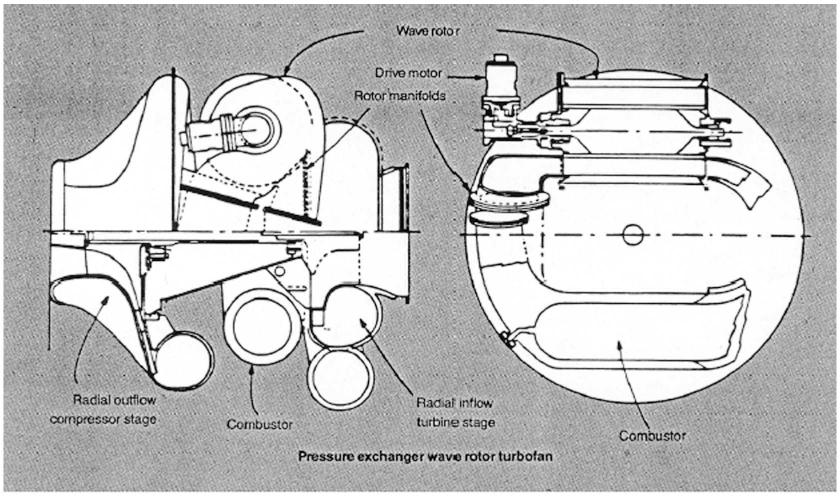


Fig. 17 Conceptual design of a turbofan engine incorporating a wave rotor, taken from Ref. [145]

troduced by Mathur based on the random choice method for solving the Euler equations [152,153]. The unconditionally stable code, called WRCOMP (wave rotor component), calculated the unsteady-flow process inside the wave rotor, inlet and outlet opening times, and other useful design parameters required for a preliminary design. The outputs from WRCOMP were used in a second program, called ENGINE, for turbofan jet engine performance calculations [154–156], and predicted significant performance improvement of a turbofan engine. Some improvements to WRCOMP code were begun [157,158], but the wave rotor research was terminated around 1986. NPS also sponsored the most comprehensive wave rotor conference in 1985 [159], which reviewed much of the history to that point.

Recent Research and Development

Since the 1985 wave rotor workshop, there has been significant progress in wave rotor research, supported by improved methods that include powerful computational capabilities allowing accurate simulation of the flow field inside the wave rotor, and modern experimental measurement and diagnostic techniques. Improvements in aerodynamic design, sealing technologies, and thermal control methods have been sought for applications that demand highly efficient performance. Numerous studies of the performance benefits of wave rotor topped gas turbines underline the motivation to obtain a quantum improvement in performance over conventional engines. Use of a wave rotor for confined combustion received renewed interest in Switzerland and the U.S., and was further stimulated by developments in related processes of pulsed detonation engines. The following recent and ongoing efforts indicate that a modern computational methods and experimental tools are allowing improved understanding, and more realism in integration with conventional machines.

NASA Glenn Research Center (1988–Present). Since the late 1980s, a sustained research program at NASA Lewis (now Glenn) Research Center (GRC), collaborating with the U.S. Army Research Laboratory (ARL) and Rolls-Royce Allison has aimed to develop and demonstrate the benefits of wave rotor technology for future aircraft propulsion systems [10]. In 1993, using a thermodynamic approach to calculate the thermal efficiency and specific power, Wilson and Paxson [160] published a feasibility study for topping jet engines with wave rotors. Applied to the case of an aircraft flying at Mach 0.8, they have shown that a wave-rotor-topped engine may gain 1–2% in efficiency and 10–16% in specific power compared to a simple jet engine with the same overall pressure ratio and turbine inlet temperature. In 1995, Welch et al. [17] predicted a 19–21% increase in specific power and a 16–17% decrease in specific fuel consumption compared with the baseline engines for small (300 to 500 kW) and intermediate (2000 to 3000 kW) wave-rotor-enhanced turboshaft engines. A wave-rotor-enhanced large turbofan engine, equal in thrust to the baseline engine, showed a 6–7% reduction in thrust specific fuel consumption. Welch has also studied the possibility of curving the channels to create a wave turbine [161,162].

Early in the program, Paxson developed a quasi-one-dimensional gasdynamic model and a computational code to calculate design geometry and off-design wave rotor performance [163,164]. The code uses an explicit, second order, Lax-Wendroff type TVD scheme based on the method of Roe to solve the unsteady-flow field in an axial passage for time-varying inlet and outlet port conditions. It employs simplified models to account for losses due to gradual passage opening and closing, viscous and heat transfer effects, leakage, flow incidence mismatch, and non-uniform port flow field mixing. In order to verify wave rotor flow predictions and to assess the effects of various loss mechanisms [165], a three-port wave-divider machine was constructed and tested [166–168] in a new wave rotor laboratory facility at GRC. Concurrently, the nonideal behavior and losses due to multi-dimensional effects were studied by Welch [169–171] and Larosiliere [172,173]. Welch has also established macroscopic and passage-averaged models to estimate the performance enhancements of wave rotors [16,174]. Based on experimental data, Paxson further improved the one-dimensional model [165,175,176] and used it to evaluate dynamic behavior, startup transients, and channel area variation [177–179]. This model was then used as a preliminary design tool to evaluate and optimize a four-port wave rotor cycle for gas turbine topping [180]. A through-flow cycle was chosen based on several perceived merits, including relatively uniform rotor temperature, and the feasibility of integration with gas turbomachinery. As a result of these studies, a new four-port wave rotor was designed and built [181] to test the performance of this concept under scaled laboratory conditions. A photograph of the NASA four-port wave rotor is shown in Fig. 18. However, a study by Rolls-Royce Allison [11] discussed further below indicated that thermal loads on the rotor and ducting predicted for the NASA wave rotor cycle in real engine conditions may be difficult to manage. In response, Nalim and Paxson [182,183] proposed an alternative cycle with a combustor bypass significantly lowering thermal loads.

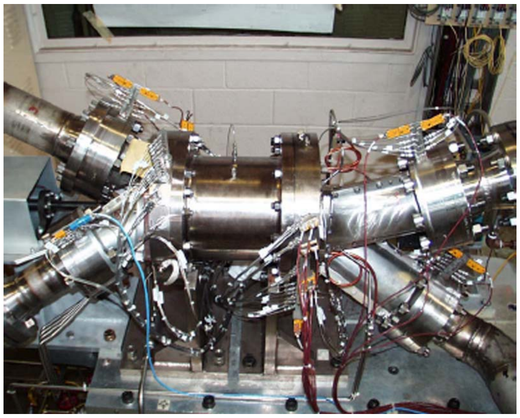


Fig. 18 NASA four-port wave rotor

In 1995, Nalim at NASA published a feasibility assessment of combustion in the channels of a wave rotor, for use as a pressure-gain combustor [184]. Combustion prediction capability was added to the wave rotor code by Nalim and Paxson [185], enabling the exploration of wave cycles involving both detonation and deflagration modes of combustion. A single reaction progress variable was used for uniform mixtures, while multiple species were represented for stratified mixtures in deflagration modes. A single-step mixing controlled reaction was combined with a simple eddy diffusivity model, ignition-temperature kinetics, and a simple total-energy based flammability limit [186]. The performance of detonative and deflagrative cycles was studied by combined computational fluid dynamics and system simulation. It was determined that deflagrative combustion with longitudinal fuel stratification could be accomplished over a reasonable time in wave rotors. The code's one-dimensional detonation prediction capability later became widely useful for study of pulse detonation engines (PDE).

Recent NASA work focused on experimental tests with special attention to sealing technology [187–189], identified as a critical challenge in high-pressure wave rotor design. The worldwide resurgence of interest in PDE technology in the 1990s has piqued interest and found synergies in wave rotor research, especially the combustion wave rotor research at NASA.

Asea Brown Boveri (1989–1994). Following the successful results of the Comprex® supercharger by BBC, in 1989 BBC/ABB commenced a three-phase research project to employ the wave rotor concept for gas turbine engines. The first phase of the project was aimed at testing a pressure-exchange wave rotor with external

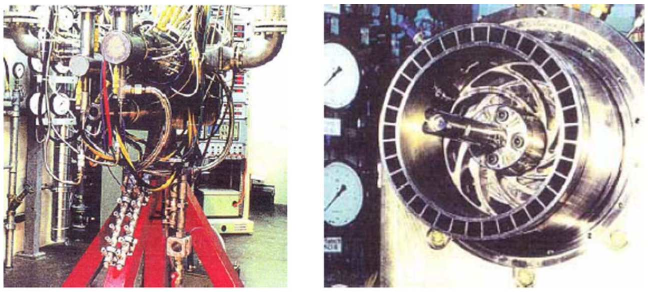


Fig. 19 ABB test rig (left) and cross section of the rotor channels (right), taken from Ref. [190]

combustion. The demonstrator engine produced 17% and 25% increases in efficiency and specific power, respectively, compared with the baseline engine [81]. The obtained significant performance improvement encouraged ABB to investigate the possibility of manufacturing an integrated combustion wave rotor, started in 1991. To explore fundamental parameters of such a constant-volume combustion device, a rotary-valved single-channel wave rotor was built and tested [82]. Using preheated and precompressed air-propane as the working fluid, combustion measurements revealed low NOx values, down to 20 ppm, at the exhaust gas due to the short residence time (1–6 ms).

After successful operation of the fixed single channel device, the design of an on combustion rotating wave rotor with 36 channels was started in 1992 [82,190]. Figure 19 shows a picture of the test rig and the cross section of the rotor channels including the rotating shaft. The inner diameter of the rotor was 20 cm and each channel had 16.5 cm length and 15×15 mm cross section. The rotor was driven by an electric motor capable of up to 5000 revolutions per minute. Both in the single-tube and 36-channel wave rot engine, spark plug and hot gas-injection self-sustaining ignition methods were utilized. Self-sustaining ignition was accomplished by employing jet injection of already-burned gas from a neighboring channel as suggested by Keller [191]. Such an ignition technique, which makes the combustion process essentially continuous without need for pulsed methods with ignition delays, has a good potential to enhance ignition reliability and burning rate, and can lead to a self-sustaining engine [192–194]. During the tests, various fuels were tried and fuel mixture was stratified via four injection nozzles. Operating in two cycles per revolution, the engine had high-pressure and low-pressure outflows per cycle, enabling it to scavenge the exhaust gas in one cycle.

The prototype engine operated successfully until the project was concluded in 1994, due to market concerns. During its operation, a number of problems were revealed and reported. They included (i) inhomogeneous mixture in the cell, resulting in a slow diffusion flame, (ii) maximum pressure reached was 9 bar, and leakage caused premature ignition and misfiring at higher chamber pressures, (iii) thermal stresses on the ignition ring, (iv) cantilever single bearing rotor support was inadequate, and (v) electromechanical device for controlling leakage gap turned out to be very complicated and sensitive. Major remedies recommended to make the system better include (a) lead away duct for leakage gas removal, (b) rotor cooling by air, (c) two-sided rotor support, and (d) mechanical control for thermal expansion.

Rolls-Royce Allison (1990–Present). Allison Engine Company (later Rolls Royce Allison, and now Rolls Royce) was a close partner in the NASA program. In 1996, Snyder and Fish [11,12] of Allison Engine Company evaluated the Allison 250 turboshaft engine as a potential platform for a wave rotor demonstration, predicting an 18–20% increase in specific power and a 15–22% decrease in specific fuel consumption. They used a detailed map of the wave rotor cycle performance accomplished by Wilson and Paxson [10,160,180]. Allison has also studied transition duct designs for integration with turbomachinery [195,196]. This was later followed by investigations of pulse detonation wave rotors in the newly formed Allison Advanced Development Company (now Rolls-Royce "LibertyWorks"). A novel four-port device was proposed [197,198] for supersonic turbofan engines [199], and was investigated in collaboration with Nalim at Indiana University Purdue University Indianapolis (IUPUI) as discussed later.

University of Florida (1992–1998). Motivated by NASA wave rotor successes, Lear et al. at the University of Florida initiated analytical and numerical methods to investigate different configurations of wave rotors. His team developed an unsteady two-dimensional numerical code using a direct boundary value method for the Euler equations to analyze the flow in wave rotors and their adjoining ducts, treating the straight or curved channel walls as constraints imposed via a body force term [200,201]. The code was used to simulate the flow fields of the three-port NASA wave rotor and the GPC wave engine [137]. They also introduced a preliminary design method for selecting the wave engine inflow and outflow blade angles, and an analytical thermodynamic model [202], which predicted potential increase in specific power of 69% and a 6.8% increase in thermal efficiency over a conventional gas turbine. A parametric study of gradual opening effects on wave rotor compression processes is also reported [203].

ONERA in France (1995–1999). Fatsis and Ribaud at the French National Aerospace Research Establishment (ONERA) have investigated wave rotor enhancement of gas turbines in auxiliary power units, turboshaft, turbojet, and turbofan engines [14,204], accounting for compression and expansion efficiency, as well as mixing and pressure losses in the ducting. Their results show the largest gains for engines with a low compressor pressure ratio and high turbine inlet temperature, such as turboshaft engines and auxiliary power units, consistent with NASA GRC studies [205]. They have also developed a one-dimensional numerical code based on an approximate Riemann solver taking into account viscous, thermal, and leakage losses [14,206], and applied it to three-port, through-flow, and reverse-flow configurations.

Recent Academic Work

Besides ongoing research mainly at NASA, Allison Advanced Development Company (AADC), and ETH Zurich, a few universities have been conducting wave rotor research. To the knowledge of the authors, the universities listed below are active in this field.

Purdue School of Engineering and Technology, IUPUI (1997–Present). Recent research at Indiana University Purdue University Indianapolis (IUPUI) by Nalim and coworkers has focused on the combustion wave rotor concept [83,84], following initial work at NASA described before. Deflagrative combustion with longitudinal fuel stratification has yielded a wave rotor geometry competitive with pressure-exchanger designs using a separate combustor [186]. Nalim has highlighted the importance of leakage flow temperature and thermal management of end-wall temperatures illustrating the impact of the hot ignition gas and the cold buffer zones on the end walls. This is consistent with the major challenges revealed by the ABB experiment [82]. Radial stratification [207] using a pre-combustion partition has been proposed to introduce a relatively cooler buffer zone close to the leakage gaps, reducing hot gas or fuel leakage to the rotor cavity. Figure 20 is a contour plot of the temperature contour from a simulation of deflagrative combustion in a stoichiometric partition region propagating into a leaner mixture in the main chamber. Above and below the partitions, there is no fuel, and gas may leak out or in without danger of overheating or pre-ignition. These thermal management approaches are possible utilizing extensive cycle design studies and analysis, and seek to alleviate the challenges previously recognized by ABB and NASA. This technique also helps burn leaner mixtures, resulting in reduced NO_x emis-

Fig. 20 Temperature distribution of partition exit flow, taken from Ref. [208]

sions, similar to other pilot combustion or lean-burn techniques in conventional engines [208]. For this approach, radial leakage flows [209] and different combustion models [210] have been studied in detail. These ideas have not yet been tested experimentally.

Detonative combustion cycles for propulsion engines have been also studied [211,212]. Interest in detonative combustion initially focused on pulsed detonation engines (PDEs) has evolved to the consideration of the wave rotor as an effective implementation of the concept [213], as the wave rotor provides automatic high-speed valving, nearly steady inflow and outflow, and the use of one or few steady ignition devices for multiple tubes. However, detonative combustion is fundamentally restricted to highly energetic mixtures and sufficiently large passage widths, and generates strong pressure waves. This results in the outflow being highly nonuniform in pressure, velocity, and possibly temperature. To better utilize the output of a wave rotor PDE, it has been proposed to add an ejector element to the wave rotor [214,215]. The rotary wave ejector admits bypass air after the detonation tubes to transfer energy and momentum. Numerical simulations using a quasi-one-dimensional code, modified to account for radial-type bypass flows, have shown that the specific impulse at static thrust conditions can be doubled, after accounting for flow-turning and shock losses, comparing with an equivalently loss-free PDE cycle. A sample wave diagram and a schematic sketch are given in Fig. 21, where the cold ejector gas flow is clearly distinguishable.

IUPUI has also investigated [198] the four-port detonation wave rotor proposed by AADC (now Rolls-Royce "Liberty-Works") [197], in which a recirculation duct allows air that is compressed by the shock of a detonation wave to be reinjected with fuel. Air-buffer regions both between the fuel/air-combusted gas interface and at the exit end plate are inherent in the cycle design, allowing self-cooling of the walls. The inflow and outflow of this engine concept is designed to be nearly uniform and acceptable to modern turbines, compared to conventional rotary detonation cycles, as shown in Fig. 22.

A computational and experimental program is currently being conducted at IUPUI in collaboration with Rolls-Royce to investigate the combustion process and performance of a wave rotor with detonative and near-detonative internal combustion. A preliminary design method based on a sequence of computational models has been developed to design wave processes for testing in an experimental test rig [216].

University of Tokyo (2000–Present). Nagashima et al. have developed one-dimensional [217] and two-dimensional [218] numerical codes to simulate the flow fields inside through-flow four-port wave rotors, including the effects of passage-to-passage leakage. The codes have been validated with experimental data obtained by a single-channel wave rotor experiment. The test rig, shown in Fig. 23, consists of a stationary single tube, and two rotating plates connected to a shaft driven by an electric motor. This group has also explored the idea of using wave rotors for ultra-micro gas turbines manufactured using microfabrication technology [219,220].

Michigan State University (2002–Present). The MSU wave rotor group has initiated studies to evaluate wave rotor technology benefits for several thermal cycle applications. Two unrecuperated microturbines (30 and 60 kW) implementing various wave-rotor-topping cycles were predicted to have overall thermal efficiency and specific work enhancement up to 34% for the smaller engine and 25% for the larger engine, using a four-port wave rotor with a compression ratio of 1.8 [221,222]. A similar approach has predicted an improvement up to 15% of overall efficiency and specific thrust in a turbojet engine using the wave-rotor-topping cycle of 30 kW microturbine [223]. Furthermore, multi-parametric performance maps for different wave rotor implementations have been generated specifying optimum operating points for both untopped baseline and topped engines [224]. Using predicted performance results, an analytical preliminary design procedure was developed for the critical high-pressure phase of four-port wave rotors [225,226]. To validate and support the accuracy of the analytical results, comparisons with numerical results of a test case was performed [227].

Recently, a unique application of wave rotors in refrigeration cycles using water (R718) as a refrigerant has also been studied [228–231]. The wave-rotor implementation can increase efficiency and reduce the size and cost of R718 units. A three-port wave rotor has been introduced as a condensing wave rotor that employs pressurized water to pressurize, desuperheat, and condense the refrigerant vapor—all in one dynamic process. In addition to the possibility of an additional rise of the vapor pressure,

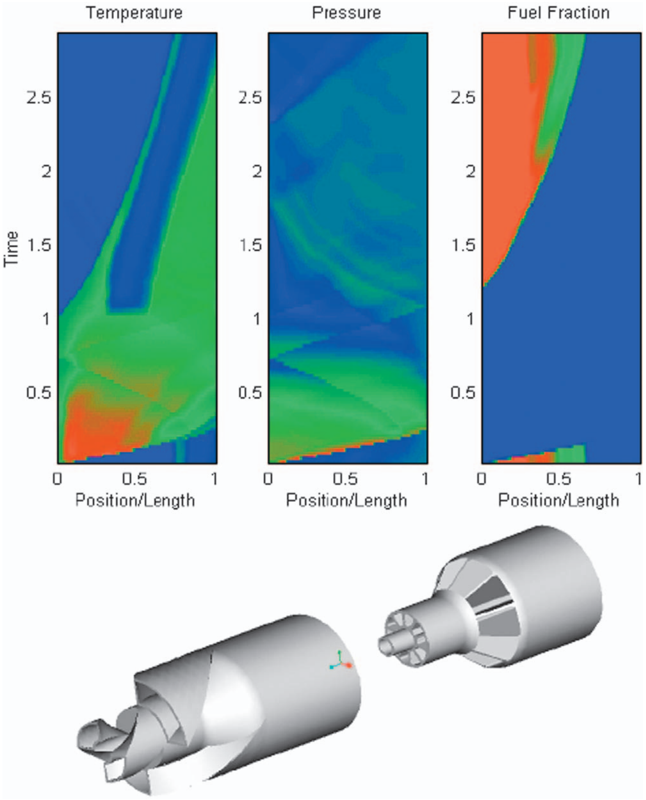


Fig. 21 Rotary Wave Ejector Pulse Detonation Engine, taken from Ref. [214]

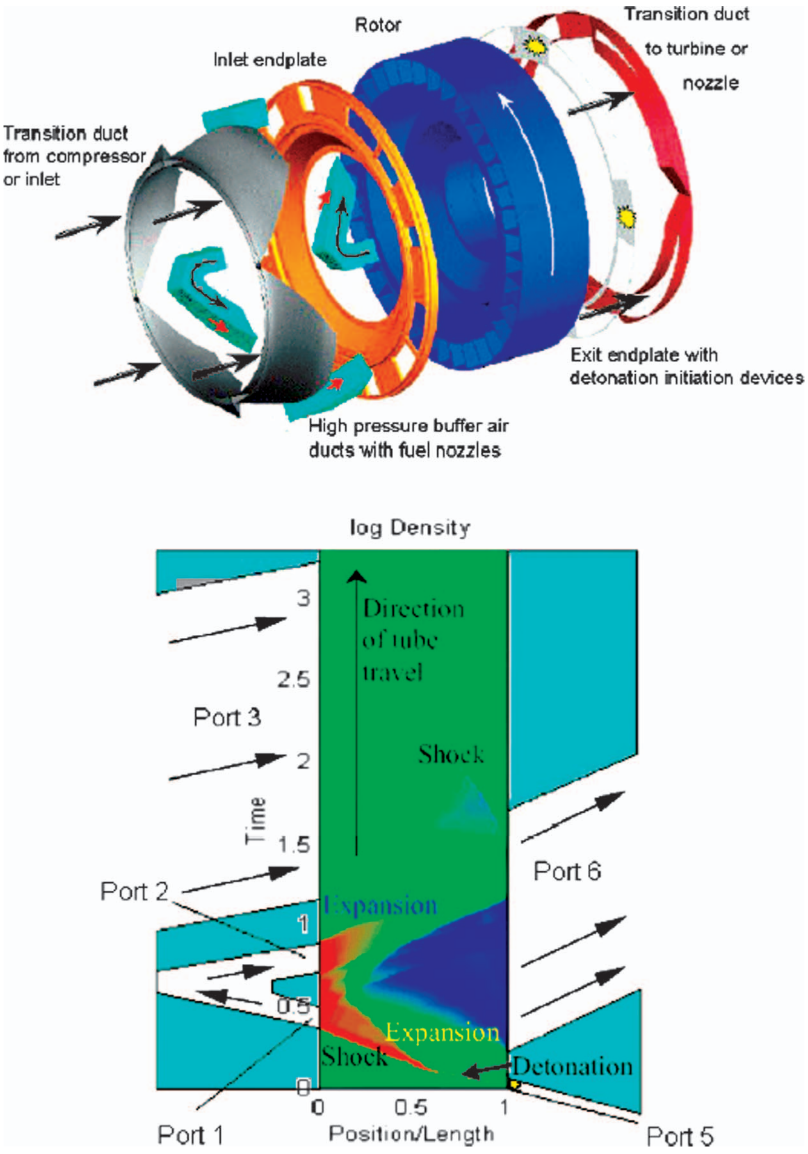


Fig. 22 Wave Rotor Pulse Detonation Engine, the "CVC" Engine, taken from Ref. [198]

the condensing wave rotor eliminates the need of a bulky condenser because full condensation occurs inside the rotor channels. Furthermore, adding a condensing wave rotor to a water refrigeration cycle allows for a lower pressure ratio of the compressor, which is crucial for R718 chiller technology. Figure 24 shows a schematic of a R718 cycle using a three-port condensing wave rotor.

Investigations of the feasibility and potential of integrating wave rotors in microfabricated gas turbines are also being pursued at MSU [3,232–235]. Ultra-micro gas turbines (UμGT) have shown difficulties in obtaining high overall thermal efficiency and output power. Utilizing wave rotor technology was suggested to improve the performance of UμGT. The wave rotor can enhance both the overall thermal efficiency and specific work output, based on MSU studies that show an efficiency of the compression between 70% and 80% can be achieved in an ultra-micro wave rotor. Several different conceptual designs for a four-port wave rotor integrated into a baseline UμGT are reported [232]. Additionally, in a collaboration between MSU and Warsaw University of Technology in Poland, the concept of radial-flow wave rotor configuration for various gas turbines applications with an emphasize on UμGT was introduced [234,235]. Figure 25 is an example of an ultra-micro wave rotor playing simultaneously the role of compressor, turbine and electric generator.

MSU and Warsaw University of Technology have also employed the commercial software FLUENT [236–239] to investigate detailed gasdynamic phenomena inside axial and radial wave rotors. FLUENT is capable of accounting for the major losses oc-

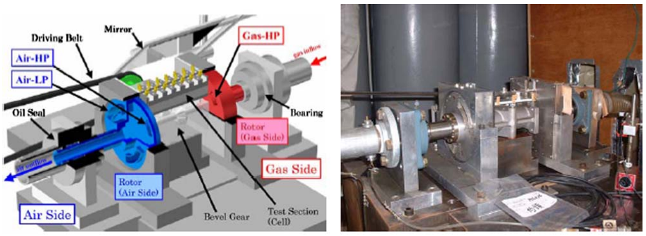


Fig. 23 University of Tokyo single-channel test rig, taken from Ref. [218] and a personal visit

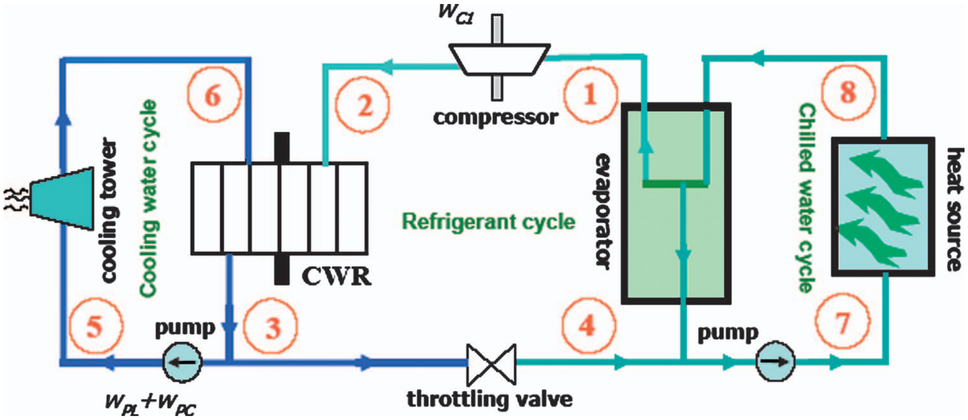


Fig. 24 Schematic of the thermodynamic model of a R718 unit enhanced by a three-port condensing wave rotor (CWR), taken from Ref. [230]

curred inside wave rotor channels including viscous and heat transfer losses, flow leakage between rotating channels, and gradual opening/closing of channels to the ports, along with other multi-dimensional features. To validate the computational results, MSU has built a wave rotor rig [240]. Piechna at Warsaw University of Technology also proposed integration of a pressure exchanger with the internal combustion wave rotor, creating an autonomous pressure wave compressor [241].

Summary

The goal of this review was to report the continued interest in wave rotor technology and its wide variety of applications. While involved researchers may have state-of-the-art and historical knowledge of this technology, it has not reached wider audiences, limiting the opportunity for meeting needs in new fields. The falling cost of computer simulation bodes well for the study of complex non-steady flow that cannot be characterized in essence by a simple global performance equation. Naturally, recent efforts were discussed in more detail as fuel costs, new research data and methods, and technology innovations have provided fresh impetus to consider the unique capabilities of wave rotor devices. Figure 26 summarizes the known history of the wave rotor research reviewed here, arranged broadly by geography. It appears that interest in gas turbine topping cycles and IC engine supercharging continues, but new opportunities appear, especially for small-scale wave machines and combustion devices that may substantially surpass conventional concepts. Continued research on sealing and thermal expansion control are needed to solve these persistent challenges, but it is important to understand and learn from the approaches of earlier workers.

Acknowledgment

The authors gratefully acknowledge several colleagues who provided valuable historical information, comments, and suggestions through personal communications. Special recognition is given to L. Guzzella at ETH in Switzerland, H. A. Nour Eldin at the University of Wuppertal in Germany, J. Piechna at Warsaw University of Technology in Poland, P. H. Snyder at Rolls-Royce North American Technologies Inc., and R. Shreeve at the Naval Postgraduate School in the U.S.

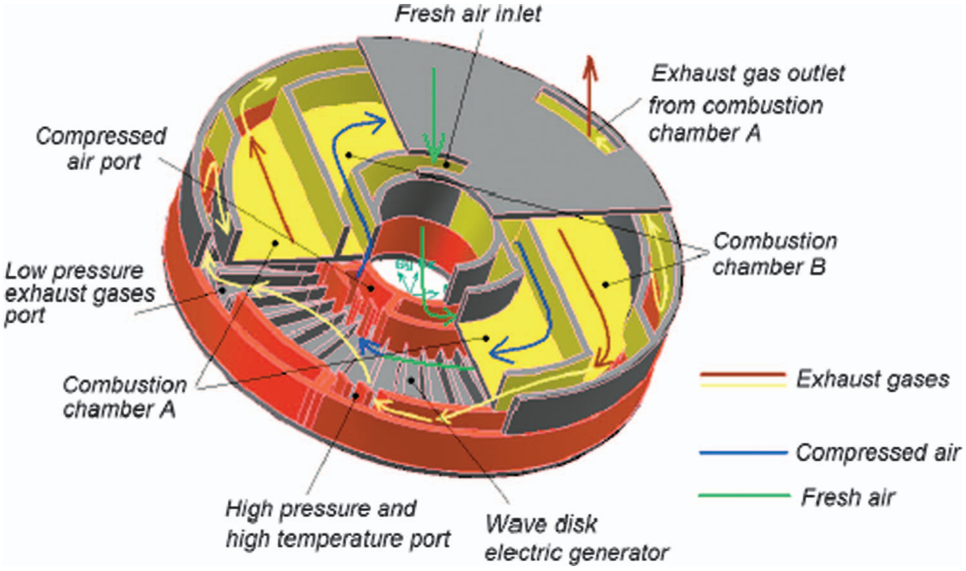


Fig. 25 Conceptual design of an ultra-micro power generator using a radial wave rotor, taken from Ref. [237]

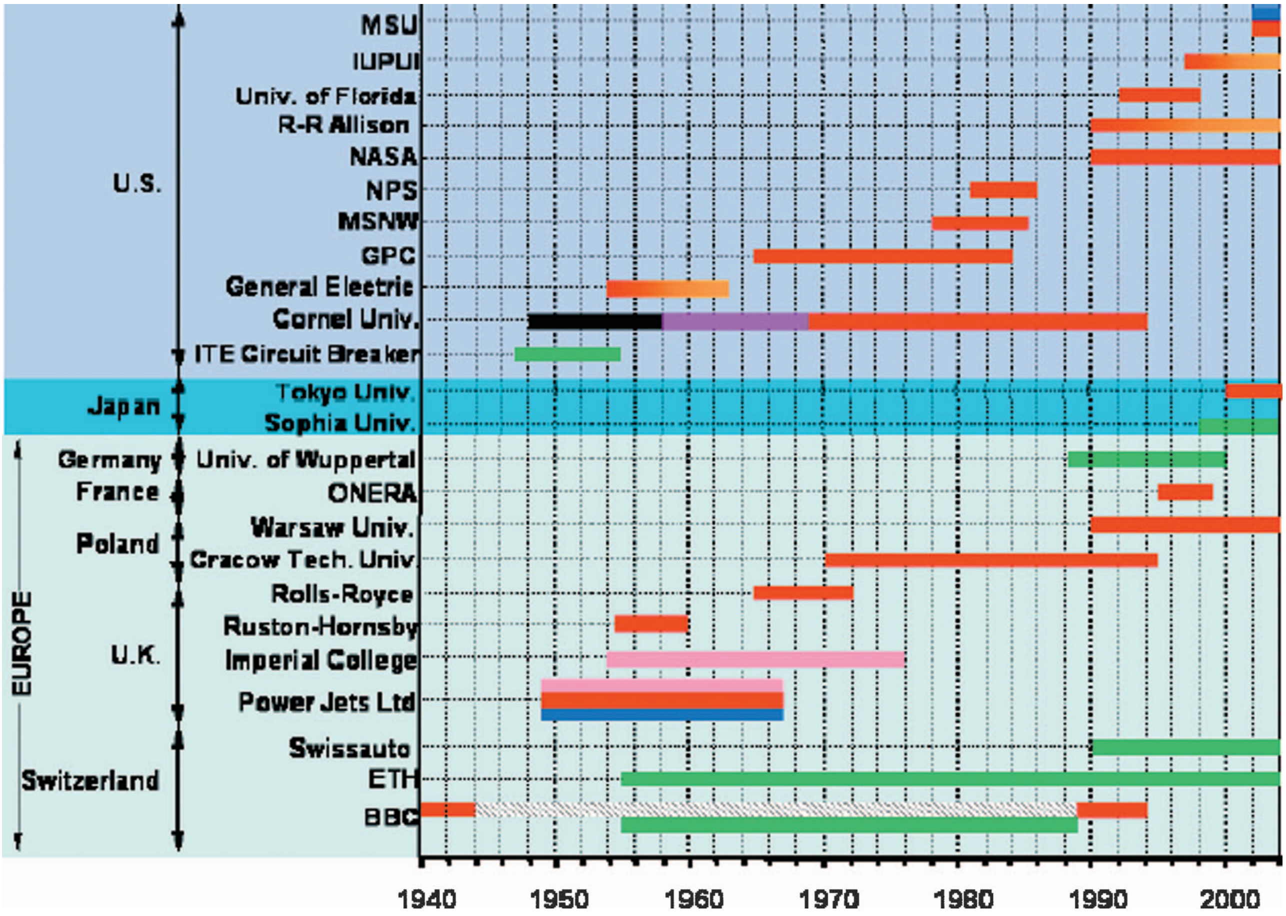


Fig. 26 Historical perspective of wave rotor technology. Red: gas turbine application. Green: IC engine supercharging. Blue: refrigeration cycle. Pink: pressure divider and equalizer. Purple: wave superheater. Orange: internal combustion wave rotors. Black: general applications.

References

[1] Weber, H. E., 1986, "Shock-Expansion Wave Engines: New Directions for Power Production," ASME Paper 86-GT-62.

[2] Weber, H. E., 1995, *Shock Wave Engine Design*, John Wiley and Sons, New York.

[3] Iancu, F., and Müller, N., 2005, "Efficiency of Shock Wave Compression in a Microchannel," Journal of Microfluid and Nanofluid, **2**(1), pp. 50–63.

[4] Paxson, D. E.,1998, "Wave Augmented Diffusers for Centrifugal Compressors," AIAA Paper 98-3401, also NASA TM-208480.

[5] Darrieus, G., 1950, "Pressure Exchange Apparatus," U.S. Patent 2526618.

[6] Kentfield, J. A. C., 1993, *Nonsteady, One-Dimensional, Internal, Compressible Flows*, Oxford University Press, Oxford.

[7] Gyarmathy, G., 1983, "How Does the Comprex Pressure-Wave Supercharger Work?," SAE Paper 830234.

[8] Zehnder, G., Mayer, A., and Mathews, L., 1989, "The Free Running Comprex®," SAE Paper 890452.

[9] Hiereth, H., 1989, "Car Tests With a Free-Running Pressure-Wave Charger—A Study for an Advanced Supercharging System," SAE Paper 890 453.

[10] Welch, G. E., 2000, "Overview of Wave-Rotor Technology for Gas Turbine Engine Topping Cycles," *Novel Aero Propulsion Systems International Symposium*, The Institution of Mechanical Engineers, London, pp. 2–17.

[11] Snyder, P. H., 1996, "Wave Rotor Demonstrator Engine Assessment," NASA CR-198496.

[12] Snyder, P. H., and Fish, R. E., 1996, "Assessment of a Wave Rotor Topped Demonstrator Gas Turbine Engine Concept," ASME Paper 96-GT-41.

[13] Berchtold, M., 1985, "The Comprex as a Topping Spool in a Gas Turbine Engine for Cruise Missile Propulsion," *Proceedings ONR/NAVAIR Wave Rotor Research and Technology Workshop*, Report NPS-67-85-008, pp. 284–290, Naval Postgraduate School, Monterey, CA.

[14] Fatsis, A., and Ribaud, Y., 1997, "Numerical Analysis of the Unsteady Flow Inside Wave Rotors Applied to Air Breathing Engines," *13th International Symposium on Airbreathing Engines*, Paper ISABE-97-7214.

[15] Mathur, A., 1985 "A Brief Review of the GE Wave Engine Program (1958-1963)," *Proceedings ONR/NAVAIR Wave Rotor Research and Technology Workshop*, Report NPS-67-85-008, pp. 171–193, Naval Postgraduate School, Monterey, CA.

[16] Welch, G. E., 1997, "Macroscopic Balance Model for Wave Rotors," J. Propul. Power, **13**(4), pp. 508–516; also AIAA Paper 96-0243, and NASA TM-107114.

[17] Welch, G. E., Jones, S. M., and Paxson, D. E., 1997, "Wave Rotor-Enhanced Gas Turbine Engines," ASME J. Eng. Gas Turbines Power, **119**(2), pp. 469–477; also AIAA Paper 95-2799, and NASA TM 106998.

[18] Knauff, R., 1906, "Converting Pressures of Liberated Gas Energy into Mechanical Work," British Patent 2818.

[19] Pearson, R. D., 1982, "Pressure Exchangers and Pressure Exchange Engines," in *The Thermodynamics and Gas Dynamics of Internal Combustion Engines* Vol. 1, R. Benson, ed., Oxford University Press, Chap. 16, pp. 903–940.

[20] Knauff, R., 1906, "Converting Internal Gas Energy into Mechanical Work," British Patent 8273.

[21] Burghard, H., 1913, British Patent 19421.

[22] Lebre, A. F., 1928, British Patent 290669.

[23] Burghard, H., 1929, German Patent 485386.

[24] Kentfield, J. A. C., 1998, "Wave Rotors and Highlights of Their Development," AIAA Paper 98-3248.

[25] Meyer, A., 1947, "Recent Developments in Gas Turbines," Journal of Mechanical Engineering, **69**(4), pp. 273–277.

[26] Real, R., 1946, "The 3000 kW Gas Turbine Locomotive Unit," Brown Boveri Rev., **33**(10), pp. 270–271.

[27] Meyer, A., 1947, "Swiss Develop New Gas Turbine Units," Electr. World, **127**, pp. 38–40.

[28] Azoury, P. H., 1992, *Engineering Applications of Unsteady Fluid Flow*, John Wiley and Sons, New York.

[29] Rose, P. H., 1979, "Potential Applications of Wave Machinery to Energy and Chemical Processes," *Proceedings of the 12th International Symposium on Shock Tubes and Waves*, Jerusalem, July 16–19, pp. 3–30.

[30] Seippel, C., 1940, Swiss Patent 225426.

[31] Seippel, C., 1942, Swiss Patent 229280.

[32] Seippel, C., 1946, "Pressure Exchanger," U.S. Patent 2399394.

[33] Seippel, C., 1949, "Gas Turbine Installation," U.S. Patent 2461186.

[34] Berchtold, M., and Gardiner, F. J., 1958, "The Comprex: A New Concept of Diesel Supercharging," ASME Paper 58-GTP-16.

[35] Berchtold, M., 1958, "The Comprex Diesel Supercharger," SAE Paper No. 63A.

[36] Berchtold, M., and Gull, H. P., 1959, "Road Performance of a Comprex Supercharged Diesel Truck," SAE Paper No. 118U.

[37] Taussig, R. T., and Hertzberg, A., 1984, "Wave Rotors for Turbomachinery," Winter Annual Meeting of the ASME, Machinery for Direct Fluid-Fluid Energy Exchange, edited by J. F. Sladky, AD-07, pp. 1–7.

[38] Doerfler, P. K., 1975, "Comprex Supercharging of Vehicle Diesel Engines," SAE Paper 750335.

[39] Schruf, G. M., and Kollbrunner, T. A., 1984, "Application and Matching of the Comprex Pressure-Wave Supercharger to Automotive Diesel Engines," SAE Paper 840133.

[40] Zehnder, G., and Mayer, A., 1984, "Comprex® Pressure-Wave Supercharging for Automotive Diesels—State-of-the-Art," SAE Paper 840132.

[41] Berchtold, M., 1985, "The Comprex®," *Proceedings ONR /NAVAIR Wave Rotor Research and Technology Workshop*, Naval Postgraduate School, Monterey, CA, Report NPS-67-85-008, pp. 50–74.

[42] Azoury, P. H., 1965-1966, "An Introduction to the Dynamic Pressure Exchanger," Proc. Inst. Mech. Eng., **180**(18), Part 1, pp. 451–480.

[43] Guzzella, L., Wenger, U., and Martin, R., 2000, "IC-Engine Downsizing and Pressure-Wave Supercharging for Fuel Economy," SAE Paper 2000-01-1019.

[44] Mayer, A., Oda, J., Kato, K., Haase, W., and Fried, R., 1989, "Extruded Ceramic—A New Technology for the Comprex® Rotor," SAE Paper 890453.

[45] Guzzella, L., and Martin, R., 1998, "The Save Engine Concept," MTZ Report 10, pp. 9–12.

[46] Burri, H., 1958, "Nonsteady Aerodynamics of the Comprex Supercharger," ASME Paper 58-GTP-15.

[47] Wunsch, A., 1971, "Fourier-Analysis Used in the Investigation of Noise Generated by Pressure Wave Machines," Brown Boveri Rev., **71**(4-5).

[48] Croes, N., 1977, "The Principle of the Pressure-Wave Machine as Used for Charging Diesel Engines," *Proceedings of the 11th International Symposium on Shock Tubes and Waves*, Seattle, July 11–14, pp. 36–55

[49] Summerauer, I., Spinnler, F., Mayer, A., and Hafner, A., 1978, "A Comparative Study of the Acceleration Performance of a Truck, Diesel Engine With Exhaust-Gas Turbocharger and With Pressure-Wave Supercharger Comprex®," The Institution of Mechanical Engineers, London, Paper C70/78.

[50] Kollbrunner, T. A., 1980, "Comprex Supercharging for Passenger Diesel Car Engines," SAE Paper 800884.

[51] Jenny, E., and Zumstein, B., 1982, "Pressure Wave Supercharger of Passenger Car Diesel Engines," The Institution of Mechanical Engineers, London, Paper C44/82.

[52] Keller, J. J., 1984, "Some Fundamentals of the Supercharger Comprex," *Presented at Winter Annual Meeting of the ASME*, Machinery for Direct Fluid-Fluid Energy Exchange, edited by J. F. Sladky, Jr., AD-07, pp. 47–54.

[53] Rebling, P., and Jaussi, F., 1984, "Field Experience with a Fleet of Test Cars Equipped with Comprex Supercharged Engines," Institution of Mechanical Engineers, London, Paper C442/84; also: SAE Paper 841308.

[54] Schneider, G., 1986, "Comprex® Pressure Wave Supercharger in an Opel Senator with 2.3 Liter Diesel Engine," Brown Boveri Rev., **73**(10), pp. 563–565.

[55] Zehnder, G., 1971, "Calculating Gas Flow in Pressure-Wave Machines," Brown Boveri Rev., **71**(4-5), pp. 172–176.

[56] Mayer, A., and Schruf, G., 1982, "Practical Experience with Pressure Wave Supercharger Comprex on Passenger Cars," The Institution of Mechanical Engineers, London, Paper C110/82.

[57] Zehnder, G., and Mayer, A., 1986, "Supercharging with Comprex to Improve the Transient Behavior of Passenger Car Diesel Engines," SAE Paper 860450.

[58] Spinnler, F., and Jaussi, F. A., 1986, "The Fully Self-Regulated Pressure Wave Supercharger Comprex for Passenger Car Diesel Engines," The Institution of Mechanical Engineers, London, Paper C124/86.

[59] Jenny, E., Moser, P., and Hansel, J., 1986, "Progress with Variable Geometry and Comprex," Institution of Mechanical Engineers Conference, London, Paper C109/86.

[60] Jenny, E., and Naguib, M., 1987, "Development of the Comprex Pressure-Wave Supercharger: In the Tradition of Thermal Turbomachinery," Brown Boveri Rev., **74**(8), pp. 416–421.

[61] Mayer, A., 1987, "The Comprex Supercharger—A Simple Machine for a Highly Complex Thermodynamic Process," Brown Boveri Rev., **74**(8), pp. 422–430.

[62] Zehnder, G., and Müller, R., 1987, "Comprex Pressure-Wave Supercharger for Car Diesel Engines," Brown Boveri Rev., **74**(8), pp. 431–437.

[63] Mayer, A., 1988, "Comprex-Supercharging Eliminates Trade-off of Performance, Fuel Economy and Emissions," SAE Paper 881152.

[64] Mayer, A., and Pauli, E., 1988, "Emissions Concept for Vehicle Diesel Engines Supercharged with Comprex," SAE Paper 880008.

[65] Mayer, A., 1988, "The Free Running Comprex—A New Concept for Pressure Wave Supercharger," SAE Document PC 55.

[66] Amstutz, A., Pauli, E., and Mayer, A., 1990, "System Optimization with Comprex Supercharging and EGR Control of Diesel Engines," SAE Paper 905097.

[67] Mayer, A., Nashar, I., Perewusnyk, J., 1990, "Comprex with Gas Pocket Control," The Institution of Mechanical Engineers, London, Paper C405/032.

[68] Mayer, A., Pauli, E., and Gygax, J., 1990, "Comprex (R) Supercharging and Emissions Reduction in Vehicles Diesel Engines," SAE Paper 900881.

[69] Barry, F. W., 1950, "Introduction to the Comprex," ASME J. Appl. Mech., March, pp. 47–53.

[70] Lutz, T. W., and Scholz, R., 1968, "Supercharging Vehicle Engines by the Comprex System," The Institution of Mechanical Engineers, London, October.

[71] Berchtold, M., and Lutz, T. W., 1974, "A New Small Power Output Gas Turbine Concept," ASME Paper 74-GT-111.

[72] Berchtold, 1974, "The Comprex Pressure Exchanger: A New Device for Thermodynamic Cycles," JSAE Paper, Tokyo.

[73] Eisele, E., Hiereth, H., and Polz, H., 1975, "Experience with Comprex Pressure Wave Supercharger on the High-Speed Passenger Car Diesel Engine," SAE Paper 750334.

[74] Groenewold, G. M., Welliver, D. R., and Kamo, R., 1977, "Performance and Sociability of Comprex Supercharged Diesel Engine," ASME Paper 77-DGP-4.

[75] Schwarzbauer, G. E., 1978, "Turbocharging of Tractor Engines with Exhaust Gas Turbochargers and the BBC-Comprex," The Institution of Mechanical Engineers, London, Paper C69/78.

[76] Walzer, P., and Rottenkolber, P., 1982, "Supercharging of Passenger Car Diesel Engines," The Institution of Mechanical Engineers, London, Paper C117/82.

[77] Wallace, F. J., and Aldis, C. A., 1982, "Comprex Supercharging Versus Turbocharging of a Large Truck Diesel Engine," The Institution of Mechanical Engineers, London, Paper C39/82.

[78] Hong-De, J., 1983, "Two-Dimensional Unsteady Flow in Comprex Rotor," Tokyo International Gas Turbine Congress, Paper 83-Tokyo-IGTC-59.

[79] Zhang, H. S., and So, R. M., 1990, "Calculation of the Material Interface in a Pressure-Wave Supercharger," Proc. Inst. Mech. Eng., Part A, **204**(A3), pp. 151–161.

[80] Hitomi, M., Yuzuriha, Y., and Tanake, K., 1989, "The Characteristics of Pressure Wave Supercharged Small Diesel Engine," SAE Paper 89054.

[81] Zauner, E., Chyou, Y. P., Walraven, F., and Althaus, R., 1993, "Gas Turbine Topping Stage Based on Energy Exchangers: Process and Performance," ASME Paper 93-GT-58.

[82] Anonymous, 1997, "A Pressure-Wave Machine with Integrated Constant-Volume Combustion," NEFF Funding of Swiss Energy Research 1977-1997, Project No. 426, pp. 142–153.

[83] Nalim, M. R., and Pekkan, K., 2003, "Internal Combustion Wave Rotors for Gas Turbine Engine Enhancement," ASME Paper IGTC-2003-FR-303.

[84] Nalim, M. R., and Pekkan, K., 2003, "A Review of Rotary Pressure-Gain Combustion Systems for Gas Turbine Applications," ASME Paper GT-2003-38349.

[85] Nour Eldin, H. A., Oberhem, H., and Schuster, U., 1987, "The Variable Grid-Method for Accurate Animation of Fast Gas Dynamics and Shock-Tube Like Problems," *Proceedings of the IMACS/IFAC International Symposium on Modeling and Simulation of Distributed Parameter Systems*, Japan, pp. 433–440.

[86] Oberhem, H., and Nour Eldin, H. A., 1990, "Fast and Distributed Algorithm for Simulation and Animation of Pressure Wave Machines," *Proceedings of the IMACS International Symposium on Mathematical and Intelligent Models in System Simulation*, Belgium, pp. 807–815.

[87] Oberhem, H., and Nour Eldin, H. A., 1991, "A Variable Grid for Accurate Animation of the Nonstationary Compressible Flow in the Pressure Wave Machine," *7th International Conference on Numerical Methods in Laminar and Turbulent Flow*, U.S.

[88] Nour Eldin, H. A., and Oberhem, H., 1993, "Accurate Animation of the Thermo-Fluidic Performance of the Pressure Wave Machine and its Balanced Material Operation," *8th International Conference on Numerical Methods in Laminar and Turbulent Flow*, UK.

[89] Markarious, S. H., Nour Eldin, H. A., and Pu, H., 1995, "Inverse Problem Approach for Unsteady Compressible Fluid-Wave Propagation in the Comprex," *9th International Conference on Numerical Methods in Laminar and Turbulent Flow*, U.S.

[90] Oberhem, H., and Nour Eldin, H. A., 1995, "Accurate Animation of the Thermo-Fluidic Performance of the Pressure-Wave Machine and its Balanced Material Operation," Int. J. Numer. Methods Heat Fluid Flow, **5**(1), pp. 63–74.

[91] Markarious, S. H., Hachicho, O. H., and Nour Eldin, H. A., 1997, "Wave-Control Modeling in the Pressure-Wave Supercharger Comprex," *10th International Conference on Numerical Methods in Laminar and Turbulent Flow*, UK.

[92] Piechna, J., 1998, "Comparison of Different Methods of Solution of Euler Equations in Application to Simulation of the Unsteady Processes in Wave Supercharger," The Archive of Mechanical Engineering, **45**(2), pp. 87–106.

[93] Piechna, J., 1998, "Numerical Simulation of the Comprex Type of Supercharger: Comparison of Two Models of Boundary Conditions," The Archive of Mechanical Engineering, **45**(3), pp. 233–250.

[94] Piechna, J., 1998, "Numerical Simulation of the Pressure Wave Supercharger - Effects of Pockets on the Comprex Supercharger Characteristics," The Archive of Mechanical Engineering, **45**(4), pp. 305–323.

[95] Piechna, J., and Lisewski, P., 1998, "Numerical Analysis of Unsteady Two-Dimensional Flow Effects in the Comprex Supercharger," The Archive of Mechanical Engineering, **45**(4), pp. 341–351.

[96] Selerowicz, W., and Piechna, J., 1999, "Comprex Type Supercharger as a Pressure-Wave Transformer Flow Characteristics," The Archive of Mechanical Engineering, **46**(1), pp. 57–77.

[97] Elloye, K. J., and Piechna, J., 1999, "Influence of the Heat Transfer on the Operation of the Pressure Wave Supercharger," The Archive of Mechanical Engineering, **46**(4), pp. 297–309.

[98] Piechna, J., 1999, "A Two-Dimensional Model of the Pressure Wave Supercharger," The Archive of Mechanical Engineering, **46**(4), pp. 331–348.

[99] Oguri, Y., Suzuki, T., Yoshida, M., and Cho, M., 2001, "Research on Adapta-

tion of Pressure Wave Supercharger (PWS) to Gasoline Engine," SAE Paper 2001-01-0368.

[100] Pfiffner, R., Weber, F., Amstutz, A., and Guzzella, L., 1997, "Modeling and Model based Control of Supercharged SI-Engines for Cars with Minimal Fuel Consumption," *Proceedings of the American Control Conference*, Vol. 1, pp. 304–308.

[101] Weber, F., and Guzzella, L., 2000, "Control Oriented Modeling of a Pressure Wave Supercharger," SAE Paper 2000-01-0567.

[102] Weber, F., Spring, P., Guzzella, L., and Onder, C., 2001, "Modeling of a Pressure Wave Supercharged SI Engine Including Dynamic EGR Effects," *3rd International Conference on Control and Diagnostics in Automotive Applications*, Italy.

[103] Weber, F., Guzzella, L., and Onder, C., 2002, "Modeling of a Pressure Wave Supercharger Including External Exhaust Gas Recirculation," Proc. Inst. Mech. Eng., Part D (J. Automob. Eng.), **216**(3), pp. 217–235

[104] Spring, P., Guzzella, L., and Onder, C., 2003, "Optimal Control Strategy for a Pressure-Wave Supercharged SI Engine," ASME Paper ICES2003-645, Austria.

[105] Icingür, Y., Hasimoglu, C., and Salman, M. S., 2003, "Effect of Comprex Supercharging on Diesel Emissions," Energy Convers. Manage., **44**, pp. 1745–1753.

[106] Weatherston, R. C., and Hertzberg, A., 1967, "The Energy Exchanger, A New Concept for High-Efficiency Gas Turbine Cycles," ASME J. Eng. Power, **89**, pp. 217–228.

[107] Hendricks, J. R., 1991, "Wave Rotor Diffusers," M.S. thesis, Cornell University, Ithaca, New York.

[108] Mathis, G. P., 1991, "Wave Enhanced Gas Turbine Engine Cycles," M.S. thesis, Cornell University, Ithaca, New York.

[109] Nalim, M. R., Moscari, J. C., and Resler, E. L., 1993, "Wave Cycle Design for NO_X Limited Wave Rotor Core Engines for High Speed Propulsion," ASME Paper 93-GT-426.

[110] Resler, E. L., Moscari, J. C., and Nalim, M. R., 1994, "Analytic Design Methods for Wave Cycles," J. Propul. Power, **10**(5), pp. 683–689; also AIAA Paper 93-2523.

[111] Mocsari, J. C., 1994, "Design of Wave Rotor Topping Cycles for Propulsion and Shaftpower," M.S. thesis, Cornell University, Ithaca, New York.

[112] Nalim, M. R., 1994, "Wave Cycle Design for Wave Rotor Engines with Limited Nitrogen-Oxide Emissions," Ph.D. thesis, Cornell University, Ithaca, New York.

[113] Nalim, M. R., and Resler, E. L., 1996, "Wave Cycle Design for Wave Rotor Gas Turbine Engines with Low NO_X Emissions," ASME J. Eng. Gas Turbines Power, **118**(3), pp. 474–480; also ASME Paper 95-GT-245.

[114] Resler, E. L., 2001, "Shock Wave Propulsion," Int. J. Chem. Kinet., **33**(12), pp. 846–852.

[115] Müller, M. A., 1954, German Patents 916607 and 924845.

[116] Foa, J. V., 1960, *Elements of Flight Propulsion*, John Wiley and Sons, New York.

[117] Taussig, R. T., 1984, "Wave Rotor Turbofan Engines for Aircraft," Winter Annual Meeting of the ASME, Machinery for Direct Fluid-Fluid Energy Exchange, edited by J. F., Sladky, paper AD-07, pp. 9–45.

[118] Kentfield, J. A. C., 1985, "The Pressure Exchanger, An Introduction Including a Review of the Work of Power Jets (R & D) Ltd.," *Proceedings ONR/NAVAIR Wave Rotor Research and Technology Workshop*, Naval Postgraduate School, Monterey, CA, Report NPS-67-85-008, pp. 9–49.

[119] Jendrassik, G., 1956, "Jet Reaction Propulsion Units Utilizing a Pressure Exchanger," U.S. Patent 2757509.

[120] Jendrassik, G., 1960, "Pressure Exchangers and Applications Thereof," U.S. Patent 2946184.

[121] Spalding, D. B., 1985, "Remarks on the Applicability of Computational Fluid Dynamics to Wave Rotor Technology," *Proceedings ONR/NAVAIR Wave Rotor Research and Technology Workshop*, Naval Postgraduate School, Monterey, CA, Report NPS-67-85-008, pp. 403–449.

[122] Jonsson, V. K., Matthews, L., and Spalding, D. B., 1973, "Numerical Solution Procedure for Calculating the Unsteady, One-Dimensional Flow of Compressible Fluid," ASME Paper 73-FE-30.

[123] Matthews, L., 1969, "An Algorithm for Unsteady Compressible One-Dimensional Fluid Flow," M.S. thesis, University of London.

[124] Azim, A., 1974, "An Investigation Into the Performance and Design of Pressure Exchangers," Ph.D. thesis, University of London.

[125] Azoury, P. H., and Hai, S. M., 1975, "Computerized Analysis of Dynamic Pressure Exchanger Scavenge Processes," Proc. Inst. Mech. Eng., **189**, pp. 149–158.

[126] Azoury, P. H., 1960, "The Dynamic Pressure Exchanger-Gas Flow in a Model Cell," Ph.D. thesis, University of London.

[127] Kentfield, J. A. C., 1963, "An Examination of the Performance of Pressure Exchanger Equalizers and Dividers," Ph.D. thesis, University of London.

[128] Kentfield, J. A. C., 1968, "An Approximation Method for Predicting the Performance of Pressure Exchangers," ASME Paper 68-WA-FE-37.

[129] Kentfield, J. A. C., 1969, "The Performance of Pressure-Exchanger Dividers and Equalizers," ASME J. Basic Eng., **91**(3), pp. 361–370.

[130] Kentfield, J. A. C., and Barnes, J. A., 1976, "The Pressure Divider: A Device for Reducing Gas-Pipe-Line Pumping-Energy Requirements," *Proceedings of 11th Intersociety Energy Conversion Engineering Conference*, pp. 636–643.

[131] Kentfield, J. A. C., 1998, "Circumferential Cell-Dividers in Wave Rotors," AIAA Paper 98-3397.

[132] Pearson, R. D., 1983, "A Pressure Exchange Engine for Burning Pyroil as the End User in a Cheap Power from Biomass System," *15th International Congress of Combustion Engines*, Paris.

[133] Pearson, R. D., 1985, "A Gas Wave-Turbine Engine Which Developed 35 HP and Performed Over a 6:1 Speed Range," *Proceedings ONR/NAVAIR Wave Rotor Research and Technology Workshop*, Naval Postgraduate School, Monterey, CA, Report NPS-67-85-008, pp. 403–449.

[134] Shreeve, R., Mathur, A., Eidelman, S., and Erwin, J., 1982, "Wave Rotor Technology Status and Research Progress Report," Naval Post-Graduate School, Monterey, CA, Report NPS-67-82-014PR.

[135] Coleman, R. R., 1984, "Wave Engine Technology Development," Final report prepared by General Power Corporation for AFWAL, Contract No. AFWAL-TR-83-2095.

[136] Coleman, R. R., 1994, "Cycle for a Three-Stage Ultrahigh Pressure Ratio Wave Turbine Engine," AIAA Paper 94-2725.

[137] Lear, W. E., 1997, "Advanced Wave Rotor, Fluid-Fluid Exchanger," Phase II report prepared by Unistry Associates for NASA, Contract No. NAS3-27647.

[138] Moritz, R., 1985, "Rolls-Royce Study of Wave Rotors (1965–1970)," *Proceedings ONR/NAVAIR Wave Rotor Research and Technology Workshop*, Naval Postgraduate School, Monterey, CA, Report NPS-67-85-008, pp. 116–124.

[139] Zumdiek, J. F., Thayer, W. J., Cassady, P. E., Taussig, R. T., Christiansen, W. H., and Hertzberg, A., 1979, "The Energy Exchanger in Advanced Power Cycle Systems," *Proceedings of the 14th Intersociety Energy Conversion Engineering Conference*, Boston.

[140] Thayer, W. J., and Taussig, R. T., 1982, "Erosion Resistance and Efficiency of Energy Exchangers," ASME Paper 82-GT-191.

[141] Zumdiek, J. F., Vaidyanathan, T. S., Klosterman, E. L., Taussig, R. T., Cassady, P. E., Thayer, W. J., and Christiansen, W. H., 1979, "The Fluid Dynamic Aspects of an Efficient Point Design Energy Exchanger," *Proceedings of the 12th International Symposium on Shock Tubes and Waves*, Jerusalem.

[142] Thayer, W. J., Vaidyanathan, T. S., and Zumdiek, J. F., 1980, "Measurements and Modeling of Energy Exchanger Flow," *Proceedings of the 14th Intersociety Energy Conversion Engineering Conference*, Seattle.

[143] Thayer, W. J., and Zumdiek, J. F., 1981, "A Comparison of Measured and Computed Energy Exchanger Performance," *Proceedings of the 13th International Symposium on Shock Tubes and Waves*, Niagara Falls.

[144] Thayer, W. J., 1985, "The MSNW Energy Exchanger Research Program," *Proceedings ONR/NAVAIR Wave Rotor Research and Technology Workshop*, Naval Postgraduate School, Monterey, CA, Report NPS-67-85-008, pp. 85–116.

[145] Taussig, R. T., 1984, "Wave Rotor Turbofan Engines for Aircraft," Mech. Eng. (Am. Soc. Mech. Eng.), **106**(11), pp. 60–68.

[146] Mathur, A., 1985, "Design and Experimental Verification of Wave Rotor Cycles," *Proceedings ONR/NAVAIR Wave Rotor Research and Technology Workshop*, Naval Postgraduate School, Monterey, CA, Report NPS-67-85-008, pp. 215–228.

[147] Eidelman, S., Mathur, A., Shreeve, R. P., and Erwin, J., 1984, "Application of Riemann Problem Solvers to Wave Machine Design," AIAA J., **22**(7), pp. 1010–1012.

[148] Eidelman, S., 1985, "The Problem of Gradual Opening in Wave Rotor Passages," J. Propul. Power, **1**(1), pp. 23–28; also published at the *19th Intersociety Energy Conversion Engineering*, San Francisco, California, 1984.

[149] Eidelman, S., 1985, "Gradual Opening of Rectangular and Skewed Wave Rotor Passages," *Proceedings ONR/NAVAIR Wave Rotor Research and Technology Workshop*, Naval Postgraduate School, Monterey, CA, Report NPS-67-85-008, pp. 229–2249.

[150] Mathur, A., Shreeve, R. P., and Eidelman, S., 1984, "Numerical Techniques for Wave Rotor Cycle Analysis," American Society of Mechanical Engineers, Fluids Engineering Division (Publication) Vol. 15, presented at the 1984 Winter Annual Meeting of the American Society of Mechanical Engineers, U.S.

[151] Eidelman, S., 1986, "Gradual Opening of Skewed Passages in Wave Rotors," J. Propul. Power, **2**(4), pp. 379–381.

[152] Mathur, A., 1985, "Wave Rotor Research: A Computer Code for Preliminary Design of Wave Diagrams," Naval Postgraduate School, Monterey, CA, Report NPS67-85-006CR.

[153] Mathur, A., and Shreeve, R. P., 1987, "Calculation of Unsteady Flow Processes in Wave Rotors," AIAA Paper 87-0011.

[154] Mathur, A., 1986, "Code Development for Turbofan Engine Cycle Performance With and Without a Wave Rotor Component," Naval Postgraduate School, Monterey, CA, Report NPS67-86-006CR.

[155] Mathur, A., 1986, "Estimation of Turbofan Engine Cycle Performance With and Without a Wave Rotor Component," Naval Postgraduate School, Monterey, CA, Report NPS67-86-008CR.

[156] Roberts, J. W., 1990, "Further Calculations of the Performance of Turbofan Engines Incorporating a Wave Rotor," M.S. thesis, Naval Postgraduate School, CA.

[157] Salacka, T. F., 1985, "Review, Implementation and Test of the QAZID Computational Method with a View to Wave Rotor Applications," M.S. thesis, Naval Postgraduate School, CA.

[158] Johnston, D. T., 1987, "Further Development of a One-Dimensional Unsteady Euler Code for Wave Rotor Applications," M.S. thesis, Naval Postgraduate School, CA.

[159] Shreeve, R. P., and Mathur, A., 1985, *Proceedings ONR/NAVAIR Wave Rotor Research and Technology Workshop*, Naval Postgraduate School, Monterey, CA, Report NPS-67-85-008.

[160] Wilson, J., and Paxson, D. E., 1993, "Jet Engine Performance Enhancement

Through Use of a Wave-Rotor Topping Cycle," NASA TM-4486.

[161] Welch, G. E., 1997, "Wave Engine Topping Cycle Assessment," AIAA Paper 97-0707; also NASA TM-107371.

[162] Welch, G. E., and Paxson, D. E., 1998, "Wave Turbine Analysis Tool Development," AIAA Paper 98-3402; also NASA TM-208485.

[163] Paxson, D. E., 1992, "A General Numerical Model for Wave-Rotor Analysis," NASA TM-105740.

[164] Paxson, D. E., 1993, "An Improved Numerical Model for Wave Rotor Design and Analysis," AIAA Paper 93-0482; also NASA TM-105915.

[165] Paxson, D. E., 1995, "Comparison Between Numerically Modeled and Experimentally Measured Wave-Rotor Loss Mechanism," J. Propul. Power, **11**(5), pp. 908–914; also AIAA Paper 93-2522 and NASA TM-106279.

[166] Wilson, J., and Fronek, D., 1993, "Initial Results from the NASA-Lewis Wave Rotor Experiment," AIAA Paper 93-2521, also NASA TM-106148.

[167] Wilson, J., 1997, "An Experiment on Losses in a Three Port Wave-Rotor," NASA CR-198508.

[168] Wilson, J., 1998, "An Experimental Determination of Loses in a Three-Port Wave Rotor," ASME J. Eng. Gas Turbines Power, **120**(4), pp. 833–842; also ASME Paper 96-GT-117, and NASA CR-198456.

[169] Welch, G. E., and Chima, R. V., 1993, "Two-Dimensional CFD Modeling of Wave Rotor Flow Dynamics," AIAA Paper 93-3318; also NASA TM-106261.

[170] Welch, G. E., 1993, "Two-Dimensional Numerical Study of Wave-Rotor Flow Dynamics," AIAA Paper 93-2525.

[171] Welch, G. E., 1997, "Two-Dimensional Computational Model for Wave Rotor Flow Dynamics," ASME J. Eng. Gas Turbines Power, **119**(4), pp. 978–985; also ASME Paper 96-GT-550, and NASA TM-107192.

[172] Larosiliere, L. M., 1995, "Wave Rotor Charging Process: Effects of Gradual Opening and Rotation," J. Propul. Power, **11**(1), pp. 178–184; also AIAA Paper 93-2526, and NASA CR-191157.

[173] Larosiliere, L. M., and Mawid, M., 1995, "Analysis of Unsteady Wave Processes in a Rotating Channel," Int. J. Numer. Methods Fluids, **21**(6), pp. 467–488; also AIAA Paper 93-2527, and NASA CR-191154.

[174] Welch, G. E., and Larosiliere, L. M., 1997, "Passage-Averaged Description of Wave Rotor Flow," AIAA Paper 97-3144; also NASA TM-107518.

[175] Paxson, D. E., and Wilson, J., 1995, "Recent Improvements to and Validation of the One Dimensional NASA Wave Rotor Model," NASA TM-106913.

[176] Paxson, D. E., 1998, "An Incidence Loss Model for Wave Rotors with Axially Aligned Passages," AIAA Paper 98-3251; also NASA TM-207923.

[177] Paxson, D. E., 1996, "Numerical Simulation of Dynamic Wave Rotor Performance," J. Propul. Power, **12**(5), pp. 949–957; also AIAA Paper 95-2800, and NASA TM-106997.

[178] Paxson, D. E., and Lindau, J. W., 1997, "Numerical Assessment of Four-Port Through-Flow Wave Rotor Cycles with Passage Height Variation," AIAA Paper 97-3142; also NASA TM-107490.

[179] Paxson, D. E., 1997, "A Numerical Investigation of the Startup Transient in a Wave Rotor," ASME J. Eng. Gas Turbines Power, **119**(3), pp. 676–682; also ASME Paper 96-GT-115, and NASA TM 107196.

[180] Wilson, J., and Paxson, D. E., 1996, "Wave Rotor Optimization for Gas Turbine Topping Cycles," J. Propul. Power, **12**(4), pp. 778–785; also SAE Paper 951411, and NASA TM-106951.

[181] Wilson, J., 1997, "Design of the NASA Lewis 4-Port Wave Rotor Experiment," AIAA Paper 97-3139; also NASA CR-202351.

[182] Paxson, D. E., and Nalim, M. R., 1999, "Modified Through-Flow Wave-Rotor Cycle with Combustor Bypass Ducts," J. Propul. Power, **15**(3), pp. 462–467; also AIAA Paper 97-3140, and NASA TM-206971.

[183] Nalim, M. R., and Paxson, D. E., 1999, "Method and Apparatus for Cold-Gas Reinjection in Through-Flow and Reverse-Flow Wave Rotors," U.S. Patent 5894719.

[184] Nalim, M. R., 1995, "Preliminary Assessment of Combustion Modes for Internal Combustion Wave Rotors," AIAA Paper 95-2801;also NASA TM-107000;also published in the 1999 ASME ASME J. Eng. Gas Turbines Power, **121**(2), pp. 265–271.

[185] Nalim, M. R., and Paxson, D. E., 1997, "A Numerical Investigation of Premixed Combustion in Wave Rotors," ASME J. Eng. Gas Turbines Power, **119**(3), pp. 668–675; also ASME Paper 96-GT-116, and NASA TM-107242.

[186] Nalim, M. R., 2000, "Longitudinally Stratified Combustion in Wave Rotors," J. Propul. Power, **16**(6), pp. 1060–1068; also AIAA Paper 97-3141, and NASA TM-107513.

[187] Hendricks, R. C., Wilson, J., Wu, T., and Flower, R., 1997, "Bidirectional Brush Seals," ASME Paper 97-GT-256; also NASA TM-107351.

[188] Hendricks, R. C., Wilson, J., Wu, T., Flower, R., and Mullen, R. L., 1997, "Bidirectional Brush Seals – Post-Test Analysis," NASA TM-107501.

[189] Hendricks, R. C., Wilson, J., Wu, T., and Flower, R., 1998, "Two-Way Brush Seals Catch a Wave," Journal of Mechanical Engineering, **120**(11), pp. 78–80.

[190] Zauner, E., and Spinnier, F., 1994, "Operational behavior of a Pressure Wave Machine with Constant Volume Combustion," ABB Technical Report.

[191] Keller, J., 1993, "Method for Preparing the Working Gas in a Gas Turbine Installation," U.S. Patent 5,197,276.

[192] Bilgin, M., Keller, J. J., and Breidenthal, R. E., 1998, "Ignition and Flame Propagation Process with Rotating Hot Jets in a Simulated Wave Engine Test Cell," AIAA Paper 98-3399.

[193] Bilgin, M., 1998, "Stationary and Rotating Hot Jet Ignition Flame Propagation In a Premixed Cell," Ph.D. thesis, University of Washington, Seattle, Washington.

[194] Akbari, P., Baronia, D., and Nalim, M. R., 2006, "Single-Tube Simulation of a Semi-Intermittent Pressure-Gain Combustor," ASME Paper GT2006-91061.

[195] Gegg, S., and Snyder, P. H., 1998, "Aerodynamic Design of a Wave Rotor to High Pressure Turbine Transition Duct," AIAA Paper 98-3249.

[196] Weber, K., and Snyder, P. H., 1998, "Wave Rotor to Turbine Transition Duct Flow Analysis," AIAA Paper 98-3250.

[197] Snyder, P. H., 2002, "Pulse Detonation Engine Wave Rotor," U.S. Patent 6449939.

[198] Snyder, P. H., Alparslan, B., and Nalim, M. R., 2002, "Gas Dynamic Analysis of the Constant Volume Combustor, A Novel Detonation Cycle," AIAA paper 2002-4069.

[199] Smith, C. F., Snyder, P. H., Emmerson, C. W., and Nalim, M. R., 2002, "Impact of the Constant Volume Combustor on a Supersonic Turbofan Engine," AIAA Paper 2002-3916.

[200] Lear, W. E., and Candler, G., 1993, "Direct Boundary Value Solution of Wave Rotor Flow Fields," AIAA Paper 93-0483.

[201] Lear, W. E., and Candler, G., 1993, "Analysis of the Accuracy of Wave Rotor Boundary Conditions Using a Novel Computational Method," AIAA Paper 93-2524.

[202] Lear, W. E., and Kielb, R. P., 1996, "The Effect of Blade Angle Design Selection on Wave-Turbine Engine Performance," ASME Paper 96-GT-259.

[203] Hoxie, S. S., Lear, W. E., and Micklow, G. J., 1998, "A CFD Study of Wave Rotor Losses Due to the Gradual Opening of Rotor Passage Inlets," AIAA Paper 98-3253.

[204] Fatsis, A., and Ribaud, Y., 1999, "Thermodynamic Analysis of Gas Turbines Topped with Wave Rotors," Aerosp. Sci. Technol., **3**(5), pp. 293–299.

[205] Jones, S. M., and Welch, G. E., 1996, "Performance Benefits for Wave Rotor-Topped Gas Turbine Engines," ASME Paper 96-GT-075.

[206] Fatsis, A., and Ribaud, Y., 1998, "Preliminary Analysis of the Flow Inside a Three-Port Wave Rotor by Means of a Numerical Model," Aerosp. Sci. Technol., **2**(5), pp. 289–300.

[207] Nalim, M. R., 2003, "Partitioned Multi-Channel Combustor," U.S. Patent 6526936.

[208] Pekkan, K., and Nalim, M. R., 2003, "Two-Dimensional Flow and NO_X Emissions In Deflagrative Internal Combustion Wave Rotor Configurations," ASME J. Eng. Gas Turbines Power, **125**(3), pp. 720–733; also ASME Paper 2002-GT-30085.

[209] Pekkan, K., and Nalim, M. R., 2002, "Control of Fuel and Hot-Gas Leakage in a Stratified Internal Combustion Wave Rotor," AIAA Paper 2002-4067.

[210] Pekkan, K., and Nalim, M. R., 2002, "On Alternative Models for Internal Combustor Wave Rotor Simulation," *Proceedings of 2002 Technical Meeting of the Central State Section of the Combustion Institute*, Knoxville.

[211] Nalim, M. R., and Jules, K., 1998, "Pulse Combustion and Wave Rotors for High-Speed Propulsion Engines," AIAA Paper 98-1614.

[212] Nalim, M. R., 2002, "Wave Rotor Detonation Engine," U.S. Patent 6460342.

[213] Kailasanath, K., 2003, "Recent Developments in the Research on Pulse Detonation Engines," AIAA J., **14**(2), pp. 145–159.

[214] Izzy, Z., and Nalim, M. R., 2001, "Rotary Ejector Enhanced Pulsed Detonation System," AIAA Paper 2001-3613.

[215] Izzy, Z., and Nalim, M. R., 2002, "Wave Fan and Rotary-Ejector Pulsed Performance Prediction," AIAA Paper 2002-4068.

[216] Snyder, P., Alparslan, B., and Nalim, M. R., 2004, "Wave Rotor Combustor Test Rig Preliminary Design," ASME Paper IMECE2004-61795.

[217] Okamoto, K., and Nagashima, T., 2003, "A Simple Numerical Approach of Micro Wave Rotor Gasdynamic Design," *16th International Symposium on Airbreathing Engines*, Paper ISABE-2003-1213.

[218] Okamoto, K., Nagashima, T., and Yamaguchi, K., 2001, "Rotor-Wall Clearance Effects upon Wave Rotor Passage Flow," *15th International Symposium on Airbreathing Engines*, Paper ISABE-2001-1222.

[219] Okamoto, K., Nagashima, T., and Yamaguchi, K., 2003, "Introductory Investigation of Micro Wave Rotor," ASME Paper IGTC03-FR-302, Japan.

[220] Okamoto, K., Nagashima, T., and Yamaguchi, K., 2005, "Design and Performance of a Micro Wave Rotor," *17th International Symposium on Airbreathing Engines*, Paper ISABE-2005-1270.

[221] Akbari, P., and Müller, N., 2003, "Performance Improvement of Small Gas Turbines Through Use of Wave Rotor Topping Cycles," *2003 International ASME/IGTI Turbo Exposition*, ASME Paper GT2003-38772.

[222] Akbari, P., Nalim, M. R., and Müller, N., 2006, "Performance Enhancement of Microturbine Engines Topped with Wave Rotors," ASME J. Eng. Gas Turbines Power, **128**(1), pp. 190–202.

[223] Akbari, P., and Müller, N., 2003, "Performance Investigation of Small Gas Turbine Engines Topped with Wave Rotors," AIAA-Paper 2003-4414.

[224] Dempsey, E., Akbari, P., Müller, N., and Nalim, M. R., 2005, "Optimum Applications of Four-Port Wave Rotors for Gas Turbines Enhancement," *17th International Symposium on Air Breathing Engines*, ISABE Paper 2005–1214.

[225] Akbari, P., and Müller, N., 2003, "Preliminary Design Procedure for Gas Turbine Topping Reverse-Flow Wave Rotors," ASME Paper IGTC03-FR-301.

[226] Akbari, P., and Müller, N., 2003, "Gas Dynamic Design Analyses of Charging Zone for Reverse-Flow Pressure Wave Superchargers," ASME Paper ICES2003-690.

[227] Akbari, P., 2004, "Performance Prediction and Preliminary Design of Wave Rotors Enhancing Gas Turbine Cycles," Ph.D. thesis, Michigan State University, E. Lansing, MI.

[228] Akbari, P., Kharazi, A. A., and Müller, N., 2003, "Utilizing Wave Rotor

Technology to Enhance the Turbo Compression in Power and Refrigeration Cycles," ASME Paper IMECE2003-44222.
[229] Kharazi, A. A., Akbari, P., and Müller, N., 2004, "An Application of Wave Rotor Technology for Performance Enhancement of R718 Refrigeration Cycles," AIAA Paper 2004-5636.
[230] Kharazi, A. A., Akbari, P., and Müller, N., 2004, "Performance Benefits of R718 Turbo-Compression Cycles Using a 3-Port Condensing Wave Rotors," ASME Paper IMECE2004-60992.
[231] Kharazi, A. A., Akbari, P., and Müller, N., 2005, "Preliminary Study of a Novel R718 Compression Refrigeration Cycle Using a 3-Port Condensing Wave Rotor," ASME J. Eng. Gas Turbines Power, **127**(3), pp. 539–544; also ASME Paper GT2004-53622.
[232] Iancu, F., Akbari, P., and Müller, N., 2004, "Feasibility Study of Integrating Four-Port Wave Rotors into Ultra-Micro Gas Turbines," AIAA Paper 2004-3581.
[233] Iancu, F., Piechna, J., Dempsey, E., and Müller, N., 2005, "Ultra-Micro Wave Rotor Investigations," *5th International Workshop on Micro Nanotechnology for Power Generation and Energy Conversion Application*, PowerMEMS, Tokyo.
[234] Iancu, F., Piechna, J., Dempsey, E., and Müller, N., 2005, "The Ultra-Micro Wave Rotor Research at Michigan State University," *2nd International Symposium on Innovative Aerial/Space Flyer Systems*, Tokyo.
[235] Iancu, F., Piechna, J., and Müller, N., 2005, "Numerical Solutions for Ultra-Micro Wave Rotors," AIAA Paper 2005-5034.
[236] Piechna, J., Akbari, P., Iancu, F., and Müller, N., 2004, "Radial-Flow Wave Rotor Concepts, Unconventional Designs and Applications," ASME Paper IMECE2004-59022.
[237] Piechna, J., 2005, "The Micro Jet Wave Engine Idea," *2nd International Symposium on Innovative Aerial/Space Flyer Systems*, Tokyo.
[238] Frąckowiak, M., Iancu, F., Potrzebowski, A., Akbari, P., Müller, N., and Piechna, J., 2004, "Numerical Simulation of Unsteady-Flow Processes in Wave Rotors," ASME Paper IMECE2004-60973.
[239] Iancu, F., 2005, "Integration of a Wave Rotor to an Ultra-Micro Gas Turbine (UμGT)," Ph.D. thesis, Michigan State University, E. Lansing, MI.
[240] Akbari, P., and Müller, N., 2005, "Wave Rotor Research Program at Michigan State University," AIAA Paper 2005-3844.
[241] Piechna, J., 2003, "Autonomous Pressure Wave Compressor Device," International Gas Turbine Congress, ASME Paper IGTC03-FR-305, Japan.

Thermomechanical Design of a Heat Exchanger for a Recuperative Aeroengine

Harald Schoenenborn
e-mail: harald.schoenenborn@muc.mtu.de

Ernst Ebert

Burkhard Simon

Paul Storm

MTU Aero Engines,
Dachauer Str. 665,
Muenchen 80995, Germany

Within the framework programs of the EU for Efficient and Environmentally Friendly Aero-Engines (EEFEA) MTU has developed a highly efficient cross-counter flow heat exchanger for the application in intercooled recuperated aeroengines. This very compact recuperator is based on the profile tube matrix arrangement invented by MTU and one of its outstanding features is the high resistance to thermal gradients. In this paper the combined thermomechanical design of the recuperator is presented. State-of-the-art calculation procedures for heat transfer and stress analysis are combined in order to perform a reliable life prediction of the recuperator. The thermal analysis is based upon a 3D parametric finite element model generation. A program has been generated, which allows the automatic generation of both the material mesh and the boundary conditions. Assumptions concerning the boundary conditions are presented as well as steady state and transient temperature results. The stress analysis is performed with a FEM code using essentially the same computational grid as the thermal analysis. With the static temperature fields the static loading of the profile tubes is determined. From transient thermal calculations successive 3D temperature fields are obtained which enable the determination of creep life and LCF life of the part. Finally, vibration analysis is performed in order to estimate the vibration stress of the profile tubes during engine operation. Together with the static stress a Goodman diagram can be constructed. The combined analysis shows the high life potential of the recuperator, which is important for economic operation of a recuperative aero-engine. [DOI: 10.1115/1.1850510]

Introduction

Current turbofan aero engines have reached a very high technology level. In order to further decrease fuel consumption and emission significantly new concepts are under investigation. MTU has developed the concept of an *I*ntercooled *R*ecuperative *A*eroengine (IRA), which is shown in Fig. 1. A more detailed description can be found in Refs. [1–3]. Between the LPC and HPC the air flows through an intercooler in the bypass duct. After the HPC the air is fed into the recuperator. The combination of these two components guarantees a high potential of fuel savings up to 20% against current aero-engines over a large speed range. Due to the weight constraints of the new components, the thermomechanical design of the recuperator is very important for the overall gain in efficiency of the complete engine.

The basic design of the heat exchanger was developed by MTU some time ago, also for land based applications [4–9]. Special profiled tubes with a low aerodynamic loss and high heat transfer coefficients are used. Some new research is ongoing in order to further study the thermodynamics and aerodynamics of the matrix arrangement [10]. The heat exchanger consists of two manifold tubes (Fig. 2). The flow from the HPC enters the upper tube (distributor) from both sides and is distributed into the 3438 U-shaped profile tubes which are brazed into the manifold tubes and form the core of the matrix. The lower manifold tube (collector tube) collects the preheated air and leads it back to the combustion chamber. The hot exhaust gas from the low pressure turbine flows upwards through the matrix and is cooled down while heating up the air inside the profile tubes.

The matrix consists of 256 rows of profile tubes (Fig. 3). The profile tubes are folded from sheet metal and welded at their mating faces. Then the profile tubes are bent into the different U-shapes. Each profile tube is part of a tube set composed of either 4 or 3 profile tubes (see Fig. 3). A total of 3438 profile tubes are required for the heat exchanger. All parts of the heat exchanger are made of ALLOY625 except the wire-spacing and cushion wire-netting where INCO600 material is used. While the heat exchanger is in service the cushion wire-netting (woven metal fiber strips) damps the vibration and maintains the space between the profile tubes. To hold the system together wire-spacing is thread through the end of the cushion wire-netting. The U-shape of the matrix makes the recuperator resistant against thermal gradients. A cover of sheet metal is placed around the matrix in order to stabilize the package and add some more damping to the system.

Thermal Analysis

Model. The thermal analysis task has concentrated first on the development of a finite element model of the matrix and second on both the prediction of steady state and transient temperature analysis. The finite element model has been developed in MSC/Patran. Rather than generating a single computational model of the hex geometry, it was decided to introduce flexibility into the model generation, so that future changes in the hex geometry could be easily facilitated in the thermal modeling. A program was written in the MSC programming language, PCL, which automatically generates a finite element model of the hex based on approximately 40 input parameters. These input parameters define the complete hex geometry, including overall dimensions, profile dimensions and spacing, as well as mesh parameters defining the fineness of the FE-mesh. In addition to the geometric parameters, some 15 parameters define the character of the flow, both on the air and gas sides of the hex, to facilitate the automatic generation of the advective and convective boundary conditions for the finite element model. More work was put into the development of this program than the generation of the single thermal finite element

---

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received October 1, 2003; final manuscript received March 1, 2004. IGTI Review Chair: A. J. Strazisar. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Vienna, Austria, June 13–17 2004, Paper No. 2004-GT-53696.

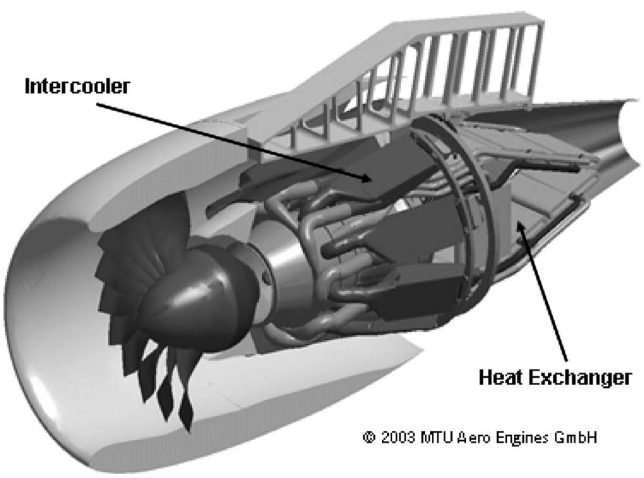


Fig. 1 IRA—intercooled recuperative aeroengine concept

model would have required; however, the benefits of this additional effort are clear. Model generation can be accomplished quickly, the quality of the mesh is independent of the user, the program is extremely flexible in terms of the range of heat exchanger geometries which can be handled, and perhaps most importantly, future changes in the hex geometry can be rapidly accommodated.

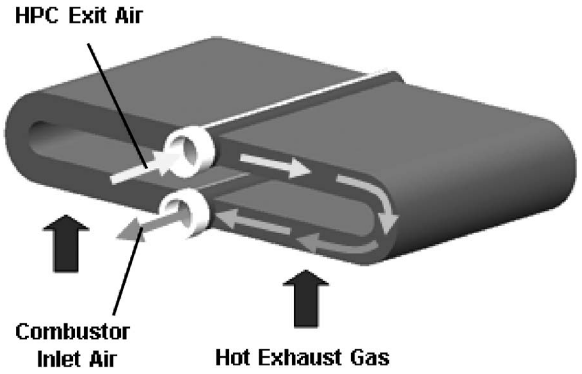


Fig. 2 Recuperator

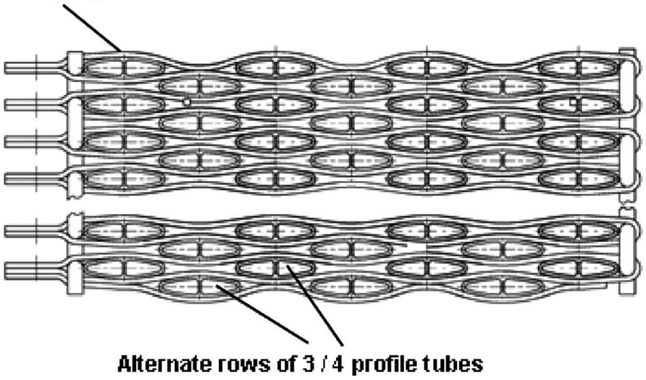


Fig. 3 Profile tube matrix cross section

The finite element model generated by the program is shown in Fig. 4. Due to symmetry conditions, the model need only consist of a single slice of the hex matrix, sliced through a row of three profile tubes on one side and four profile tubes on the other. Similarly due to symmetry conditions, only one side of the matrix needs to be modeled. Hence only half of the distributor and collector manifold walls are included in the model. Despite these simplifications arising from symmetry conditions, the model consists of approximately 51000 material nodes and 33000 material hex-8 elements.

Boundary Conditions. Boundary conditions for the thermal analysis consist of fluid (air side and hot gas side) temperatures, mass flows and heat transfer coefficients. The advective flow network facilitates the determination of the local temperature boundary conditions. Inlet temperatures are specified at the two fluid nodes only, namely for air and gas inlet temperatures. With specified mass flow distributions, the temperature at all points in the advective systems are determined from an energy balance. Hence the temperature distributions in the advective systems become coupled to the material temperature distribution and are solved simultaneously.

In terms of the mass flow distributions, the following assumptions have been made: (1) the external hot gas flow is uniformly distributed along the length of the matrix, and (2) the mass flow on the air side from the distributor manifold into the profile tubes is taken to be inversely proportional to the length of the profile tube. The advective system is shown in Fig. 5. It consists of 2900 advective elements and 4500 fluid nodes.

The inlet air and gas temperatures are based on the results of a 1D performance calulation of the aeroengine. The performance data at the recuperator inlet are given in Table I for the four principal load cases of the design mission. Typical deterioration increments, derived from a similar-sized non-intercooled, nonrecuperated engine, are added to the inlet temperatures to simulate realistic operational conditions.

The heat transfer coefficients on the air side are determined by correlations for turbulent flow in ducts. For the gas side correlations for flow over tube bundles are used. The external surface of the profile tubes is divided into four regions, for which the heat transfer coefficients are determined. The regions are shown in Fig. 6. This figure also shows the interaction between the convective, the advective and the material elements.

Radiation is not accounted for in the model; however, at the temperature levels and temperature distribution encountered in the hex matrix, radiative heat transfer may be neglected.

The entire mission profile is conservatively represented for purposes of a LCF life analysis by a minicycle, the first part representing the initial acceleration from idle to take-off conditions and climb to altitude, and the second part representing deceleration.

In undertaking a transient thermal analysis, the transient behavior of the boundary conditions, such as hex inlet temperatures and mass flows, during such acceleration and deceleration maneuvers must be known. The transient behavior of the boundary conditions was assumed from a conventional aero-engine of the same thrust class, but without intercooling or recuperation, since no data from an IRA engine had been available. One would expect, as a result of the additional thermal inertia of an intercooler and recuperator, that the IRA would have a slower thermal transient response at the exhaust exit than a conventional engine of the same size. Hence the use of transient data from a conventional engine is justified as it would lead to conservative estimates of the LCF life. The transient behavior of the hex boundary conditions for the acceleration idle-takeoff, in the form of nondimensional curves, is shown in Fig. 7 qualitatively. Similar curves exist for the deceleration from takeoff to ground idle, but are not reproduced in this paper.

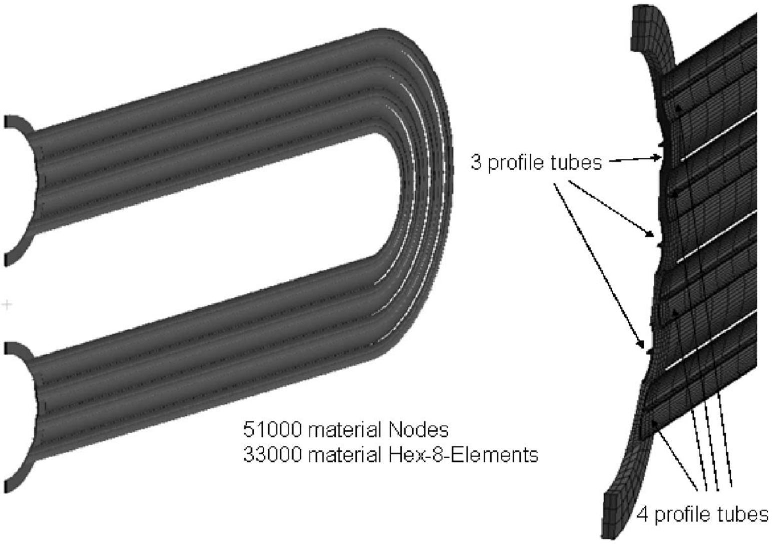


Fig. 4 Thermal FE mesh

Computational Results. Temperature calculations have been performed for different steady-state performance points such as "end take-off" using the commercial program MSC/Thermal as solver. This program allows to incorporate user defined correlation, which are used for determination of the heat transfer coefficients, during an analysis run. The material temperature distribution at this point is shown in Fig. 8. Material temperatures in the hex are far from uniform, varying considerably not only along the length of the profile tubes, but also across the tube bundle. The peak temperature is found not surprisingly on the leading edge of the first profile tube, close to the collector manifold.

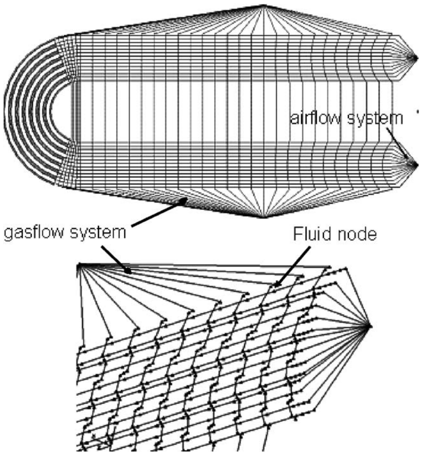


Fig. 5 Advective flow network of the thermal model

The transient temperature distribution was calculated for a minicycle "idle-takeoff-idle." Transient temperature results are plotted in Fig. 9 for selected locations on the profile tubes. The shown material temperatures are selected close to the stagnation point of the profiles. As can be seen the heat up of the profile tubes at different locations is relatively homogeneous during the acceleration phase whilst the response curves during deceleration vary significantly. The temperature curves for the locations 1F and 7F cross one another during the first few seconds of the deceleration due to the latent heat of the profile tubes. While the tubes at the location 1F are exposed directly to the lower gas temperatures, the tubes at location 7F experience a delayed gas temperature reduction due to the above mentioned mechanism. A similar behavior is seen in the transients for locations 8F and 15F, but is not strong enough to cause a crossing of the temperature curves.

Since the first profile tube close to the collector manifold experiences the highest material temperatures and the fastest response curves during a transient change of the power setting, the region marked with "1F" in Fig. 9 will be the most critical one of the hex tubes in terms of the matrix life.

Structure Mechanical Analysis

The profile tubes with their thin wall thickness are the life determining parts of the recuperator. Thus, some effort has been made in order to perform an analytical life prediction of this part as up to now only little practical experience has been collected. ABAQUS was used as FEM code.

Table 1 Recuperator inlet conditions for the four primary performance points of the thermo-mechanical analysis

Side	Property	Unit	End take-off	Max climb	Avg. cruise	Reverse
Air	Mass flow	[kg/s]	6.83	2.39	1.7	4.16
	Inlet pressure	[kPa]	3091	1084	760	1826
	Inlet temperature	[K]	722	597	507	608
Gas	Mass flow	[kg/s]	9.16	3.22	2.29	5.56
	Inlet pressure	[kPa]	166	55	39	125
	Inlet temperature	[K]	944	917	888	920

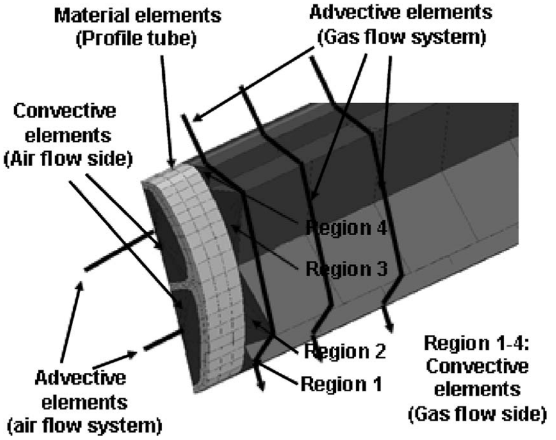


Fig. 6 Interaction between advective flow network and material mesh at a profile tube

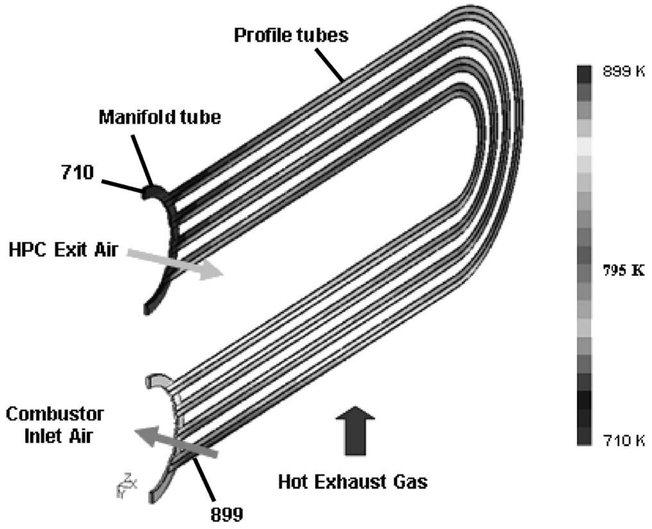


Fig. 8 Temperature distribution at takeoff conditions

Model. In Fig. 10 a finite-element model of the matrix is shown. Due to symmetry conditions the model is reduced to one row with four profile tubes. Nevertheless, this model contains more than 150 000 nodes and 28 000 twenty-noded hexaeder elements. The damping wires are modeled as multipoint constraints between the profile tubes. The boundary conditions at the symmetry plane of the manifold tubes were varied according to the type of analysis.

Static Stress. Figure 11 presents the stress results of an analysis with thermal load and internal pressure, which is applied as a distributed load onto the inner surface of the profile tubes and manifolds. Here, the static mean stress is determined, which is used later in the Goodman-diagram together with the results from the vibration analysis.

Creep. A nonlinear creep analysis was performed with a special creep user subroutine developed by MTU. A mission typical for a trans-atlantic flight was assumed. Figure 12 shows the exhaust gas temperature and the internal profile tube pressure during the mission, made dimensionless with the maximum conditions (take-off conditions). A large number of missions was considered and the time at each temperature level was summed up. For each condition of this mission, 3D temperature distributions were calculated. The complete material creep curves for different temperatures are prepared and fed into the creep subroutine. Then for each time step the creep velocity is determined for each node using the Larson-Miller-Parameter and a differential creep strain is calculated. All these creep strains are summed up for each node yielding the creep strain distribution. Figure 13 shows an example of such a result. It can be seen that only in the first row which is exposed to the hot exhaust gas at the inner small radius, where the stress due to internal pressure is highest, a considerable creep damage is accumulated. However, this creep damage did not exceed the limit of 1.0 for the assumed number of missions.

LCF. As the mission was appropriate for the creep simulation, for a complete LCF calculation too many time steps would have been necessary. Thus, as is usual only a mini-cycle as described above is considered. In Fig. 14 the Mises stress results from this mini-cycle transient analysis is presented for the maximum stress location. It can be seen that there is only a small overshoot of about 10% in Mises stress during acceleration to

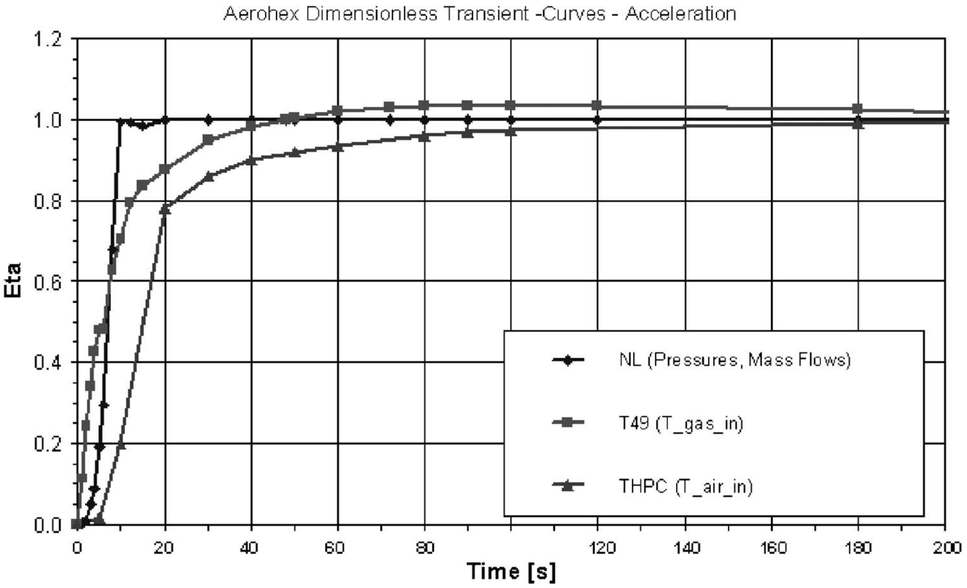


Fig. 7 Nondimensional transient behavior of boundary conditions during acceleration

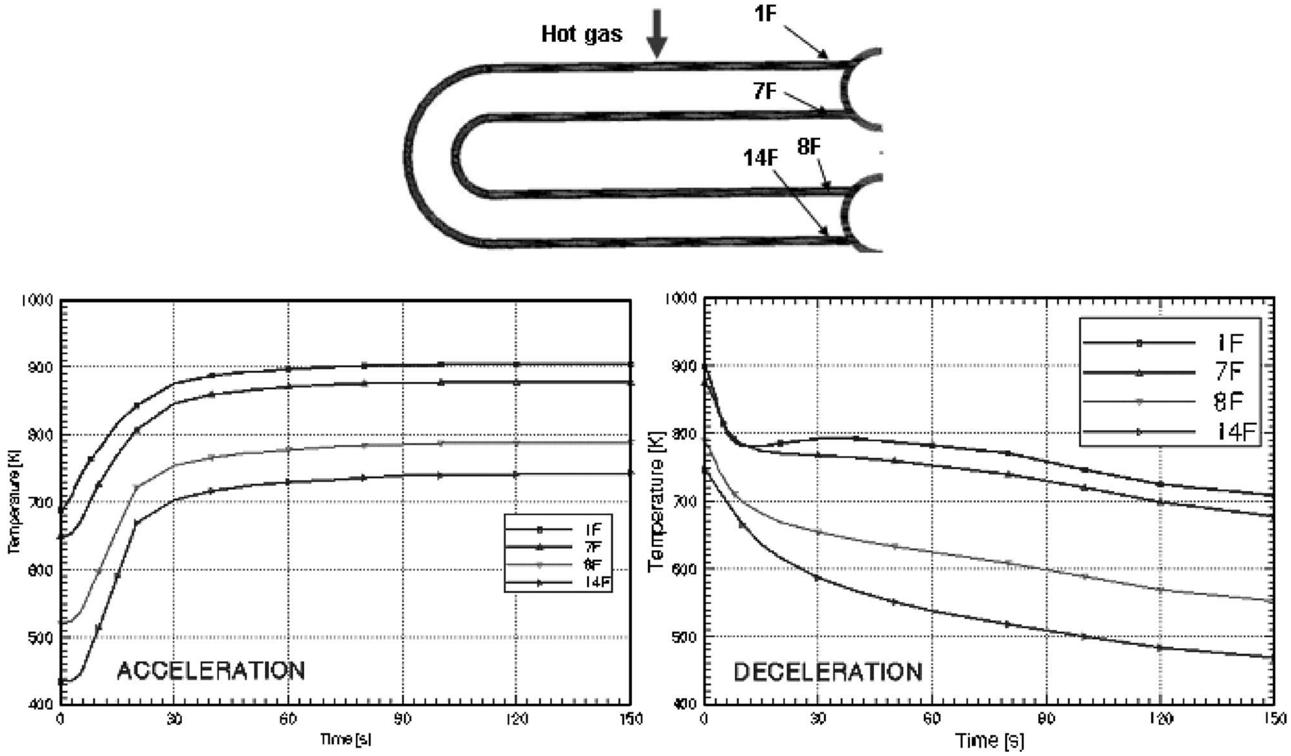


Fig. 9 Transient temperature response at selected profile tube locations during minicycle

take-off conditions against steady-state take-off conditions. The maximum principal stress did even show less overshoot. This is due to the U-shaped design of the profile tubes which result in this excellent thermomechanical behavior of the matrix.

Vibration. The basis for the vibration analysis is the determination of the eigenmodes of the profile tube package. Figure 15 shows four eigenmodes out of the many hundreds of modes. Modes 1 and 4 are the first and second bending modes in the horizontal plane, where the matrix is relatively soft. These modes are not crucial as long as the matrix is limited in this direction by the side walls in the engine. Mode 3 is the first bending mode in vertical direction and is more dangerous due to nearly no geometrical constraints in this direction. All higher modes yield vibrations of single profile tubes against each other and are not

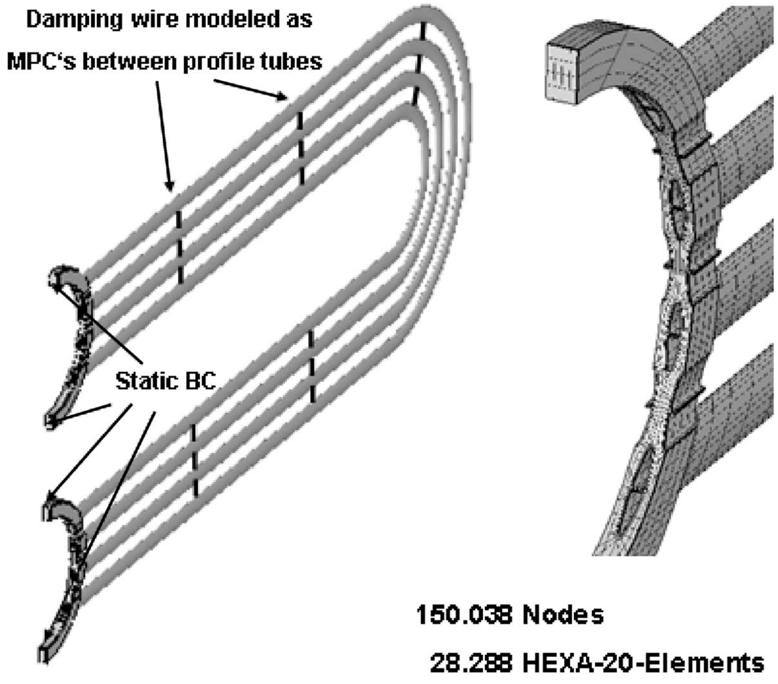


Fig. 10 FE model structure mechanics

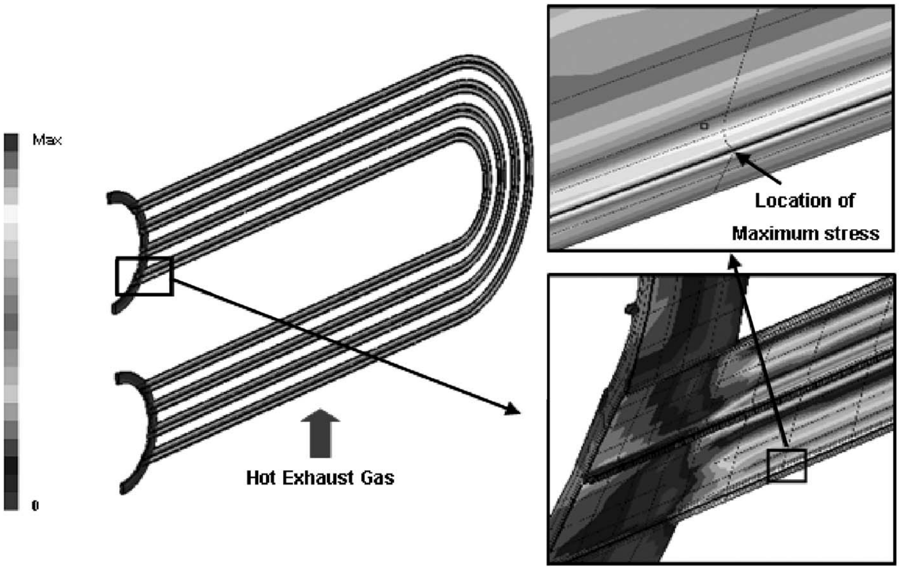


Fig. 11 Static Mises stress during takeoff

considered as critical as the matrix is highly damped. A modal-based steady-state dynamic analysis was performed in the frequency range from 5 to 2000 Hz. A uniform 1-g excitation level was used with 10% of critical damping. Figure 16 presents the Mises response stress (made dimensionless with Rm) of the location of maximum stress (see Fig. 17). The highest stresses occur for the first bending mode at 80 Hz. At higher frequencies, the stress is much lower. Figure 17 shows qualitatively the vibration stress distribution at 80 Hz. Finally it is assumed that the matrix has a vibration excitation of 5g. With this vibration stress amplitude and with the static stress determined previously a point can be drawn into the Goodman diagram (Fig. 18). It can be seen that for the take-off conditions there is still some safety margin against HCF failure. The experimental determination of vibration stress of

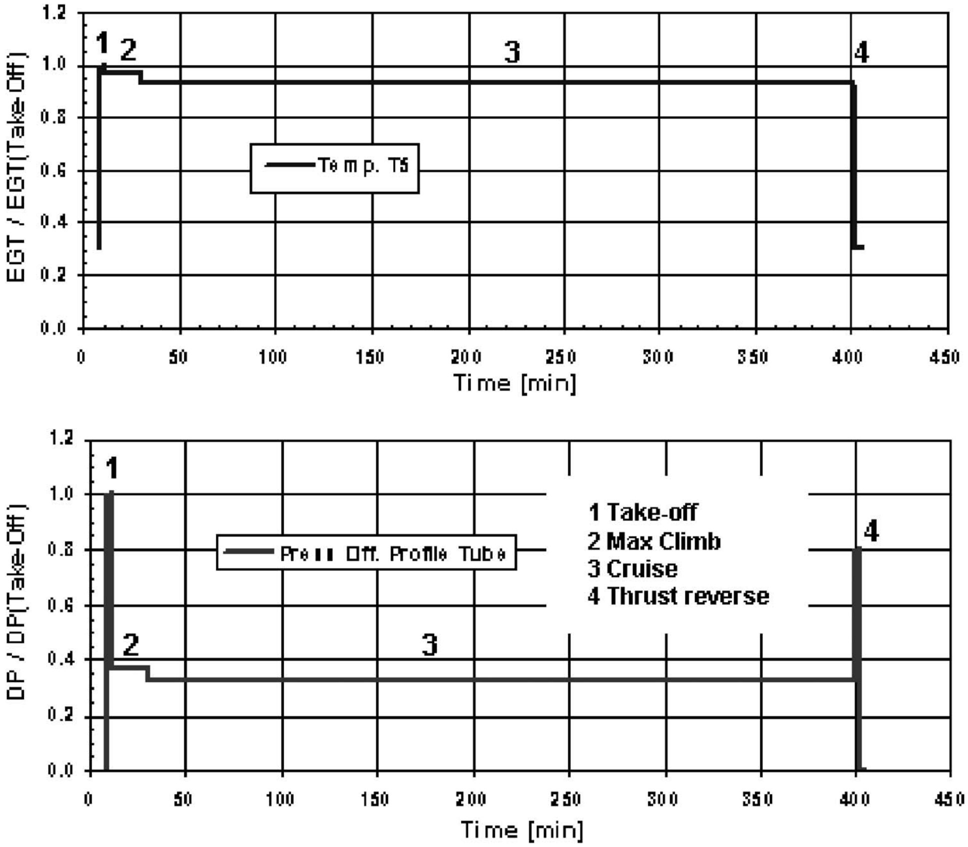


Fig. 12 Mission temperature and pressure

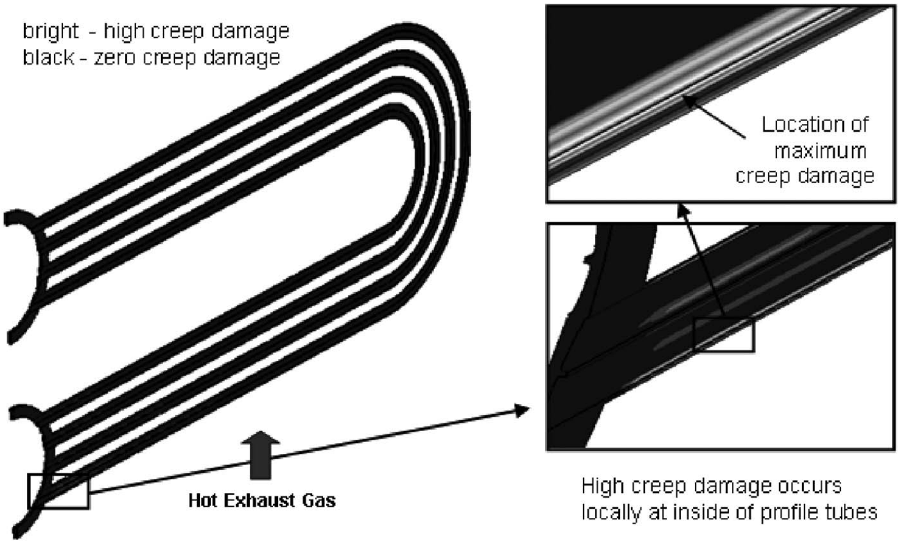


Fig. 13 Example evaluation creep calculation

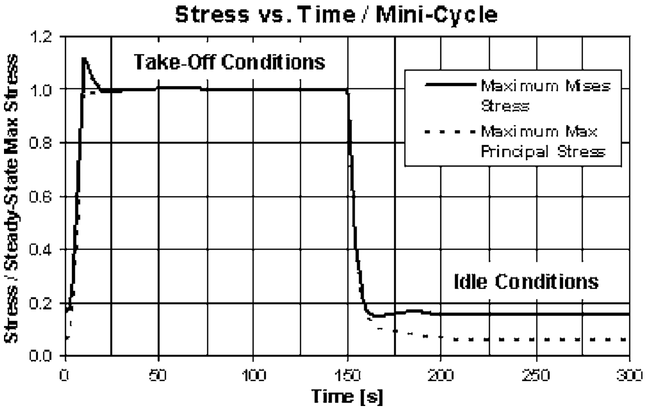


Fig. 14 TMF analysis minicycle

a matrix on a shaker table and in a real engine with strain-gauges remains a very important point in order to verify the analytical predictions.

Conclusions

The thermomechanical procedure for the analytical life prediction of a recuperator for aero engines was presented. State-of-the-art calculation methods for thermal analysis and stress analysis were used. The analytical life prediction including a combination of LCF, creep and vibration has shown that the recuperator has a life potential as usually required for aero engines. The outstanding thermomechanical design makes this type of recuperator not only attractive for aero-engines, but also for other applications in power generation. The analytical analysis shows that there are still some margins to account for difficulties which may arise in the

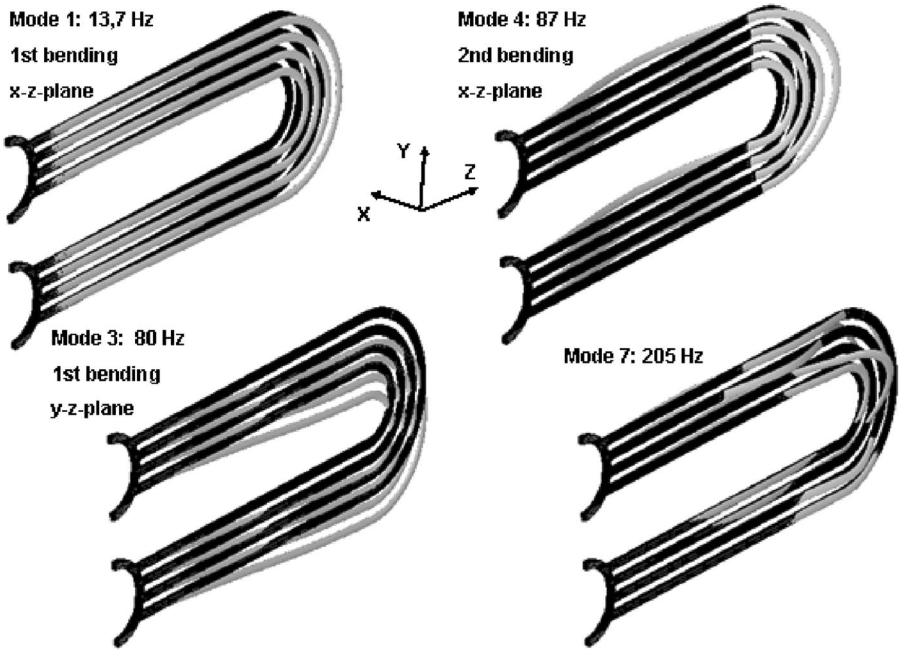


Fig. 15 Some eigenmodes

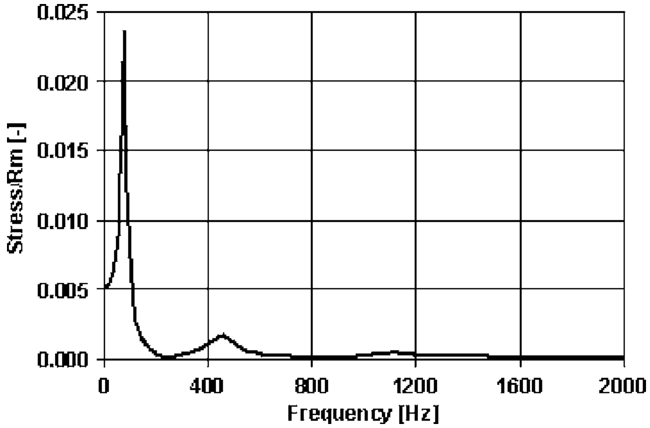


Fig. 16 Stress response at 1 g excitation at location of maximum stress (Mises)

real engine (corrosion, vibration). Now this prediction has to be proven in practical use during future component and engine test programs.

Acknowledgment

This work has been financially supported by the EU under the "Competitive and Sustainable Growth Program" Contract No. G4RD-CT-1999-00069. The authors wish to thank the EU for supporting this program. The permission for publication is gratefully acknowledged.

Nomenclature

HPC = high pressure compressor
LPC = low pressure compressor
EGT = Exhaust gas temperature
DP = Pressure difference
Rm = Ultimate tensile stress

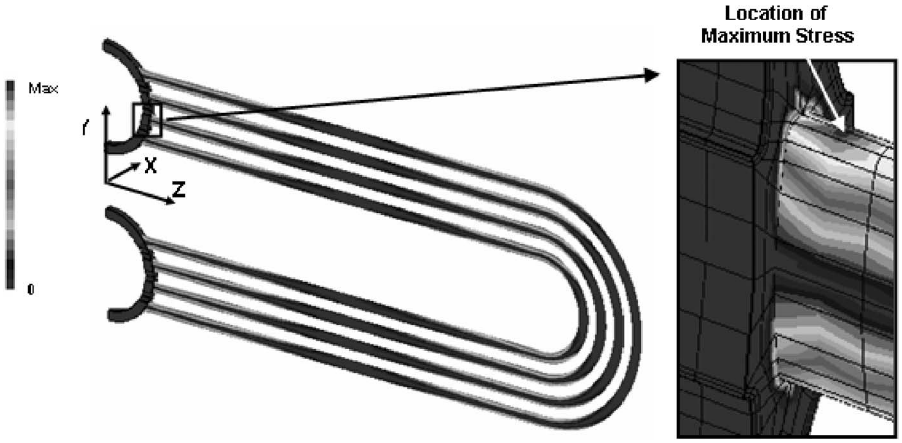


Fig. 17 Vibration stress distribution

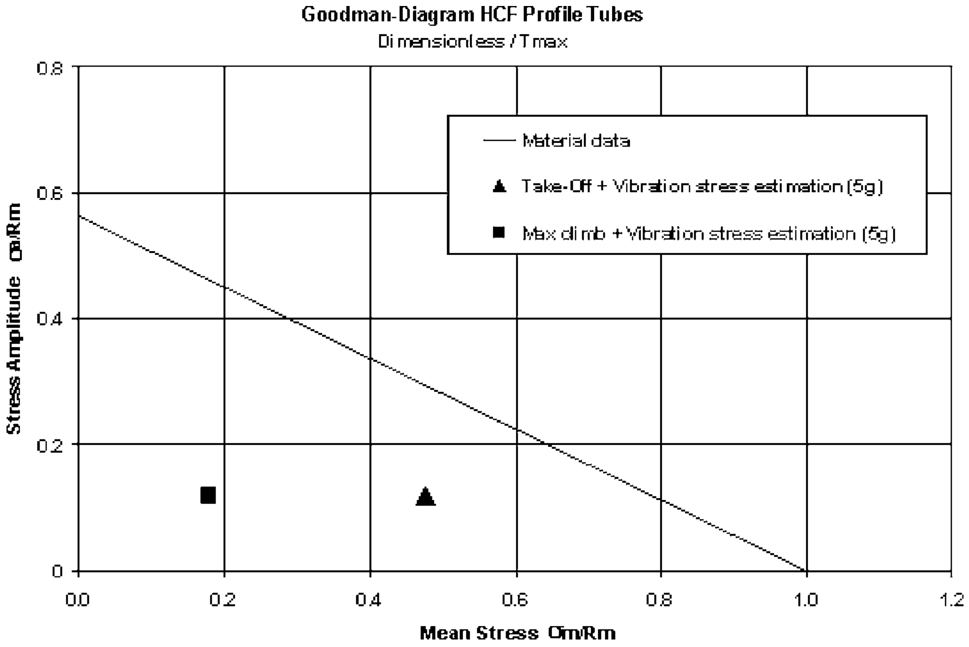


Fig. 18 Goodman diagram

References

[1] Broichhausen, K., Scheugenpflug, H., Mari, Ch., Barbot, A., 2000, "CLEAN The European Initiative Towards Ultra Low Emission Engines," ICAS 2000, Harrogate, UK.

[2] Wilfert, G., and Massé, B., 2001, "Technology Integration in a Low Emission Heat Exchanger Engine," *Proceedings of the 8th CEAS European Propulsion Forum*, Nottingham, UK.

[3] Scheugenpflug, H., Wilfert, G., and Simon, B., 2001, "Erfüllung zukünftiger Umweltanforderungen durch den Einsatz eines Wärmetauschertriebwerks," DGLR-Jahrbuch, **3**, pp. 1647–1654.

[4] Pellischek, G., and Reile, E., 1992, "Compact Energy Recovery Units for Vehicular Gas Turbines," SAE Paper 920151.

[5] Pellischek, G., and Kumpf, B., 1991, "Compact Heat Exchanger Technology for Aero Engines," ISABE Paper 91-7019.

[6] Eggebrecht, R., and Schlosser, W., 1986, "Kompakter Hochtemperatur-Wärmetauscher für Wellenleistungsturbinen," MTZ Motortechnische Zeitschrift, **47**, pp. 235–241.

[7] Duffy, R. J., and Hower, G. K., 1987, "Turbine Propulsion for Heavy Armored Vehicles," AIAA Paper 87-1911.

[8] Brockett, W., and Koschier, A. V., 1992, "LV100 AIPS Technology-for Future Army Propulsion," ASME Paper 92-GT-391.

[9] Koschier, A. V., and Mauch, H. R., 1999, "Advantages of the LV100 as a Power Producer in a Hybrid Propulsion System for Future Fighting Vehicles," ASME Paper 99-GT-416.

[10] Goulas, A., Katheder, K., Palikaras, A., and Yakinthos, K., 2003, "Flow Measurements and Investigations in a Staggered Tube Matrix of a Heat Exchanger," Int. J. of Heat & Technology, **21**(2).

Jet Engine Model for Control and Real-Time Simulations

Michael Lichtsinder

Yeshayahou Levy
e-mail: levy@aerodyne.technion.ac.il

Faculty of Aerospace Engineering,
Technion-Israel Institute of Technology,
32000 Haifa, Israel

The main objective of this paper is development of a simple real-time transient performance model for jet engine control. A jet engine arrives to its most dangerous condition during transient operation that may be triggered by fast changes of the input fuel command signal. Thus, the control system specifications are formulated to specify the maximal variance of the fuel flow command (from idle to maximum power level) during transient maneuver. Linear and piecewise-linear techniques are not always convenient and appropriate for turbine engine controller design. An alternative quasilinear simple/fast engine model is discussed in this paper. This model has maximum accuracy for maximal variance of the fuel flow input command in accordance to the jet engine control system specifications. The fast model is obtained using the Novel Generalized Describing Function, proposed for investigation of nonlinear control systems. The paper presents the Novel Generalized Describing Function definition and then discusses the application of this technique for the development a fast turbine engine simulation suitable for control and real-time applications. Simulation results are compared between the conventional and fast models and found to provide good agreement. [DOI: 10.1115/1.1915391]

1 Introduction

Fast engine models are applied both for off-line and on-line simulations. Off-line fast simulations are used for engine controller design using methods based on repeated simulations (Fuzzy Logic, Neural Networks, etc.). The on-line fast simulations are applied for:

(a) Estimation of nonmeasurable variables (such as “stall margin,” “equivalence ratio”);
(b) Backup when sensors are faulty (such as rotational speed or turbine exhaust temperature);
(c) Model based engine control.

The aerothermal nonlinear model is the most accurate to simulate engine performance and thermodynamic parameters during transient operation [1–4]. Accurate simulations of turbine engine transient characteristics are performed by adequately modeling power balance, unsteady flow volume dynamics (continuity and energy storage at engine intake, combustor diffuser, combustor chamber, exhaust diffuser and jet pipe), heat soakage, time delay in the combustion process [2,3], turbine cooling, control temperature sensor response [4], and other dynamic terms. The other model type using nonlinear differential equations of unsteady power balance including rotor moment of inertia [2–4] and algebraic aerothermal equations is described in [5–8]. These are faster and more suited for flight simulations and control, where lower accuracy is permissible. However, relatively long computation time is one of the disadvantages of this engine model type. The reason for this is the computational time required to calculate the nonlinear algebraic equations solution at each integration step during numerical solution of the differential equations.

Real-time models using *linearization* of nonlinear differential and algebraic equations are simpler and faster [9–12]. The state space or transfer function representation is used for transient performance in this linear model type. For example, one input variable (fuel flow) and four state variables (gas turbine speed, power turbine speed, rotor torque, rotor speed) are used in the linear state variable model of the T406 two-spool turbo-shaft engine [11,12]. In order to generate a linear model, the nonlinear engine model is approximated around an operating point. Partial derivatives of the nonlinear functions, or time constants and gains of the transfer functions, are obtained from the aerothermal models or test data around this operating point. Linear Time Invariant (LTI) models are not accurate for all flight conditions. However, these are widely used for engine control [12].

The other type, referred to as piecewise-linear (or linked) models, generates linear models at multiple operating points throughout the engine operating envelop. The linear models are then linked together via an interpolation scheme based upon some selected scheduling parameters such as $\dot{m}_f$, N, M, and H. This technique allows the model’s transition between intermediate points as the simulation goes through a transient. The piecewise-linear modeling technique effectively allows a simplified nonlinear model to be constructed that maintains good agreement with the original nonlinear simulation with greatly increased simulation speed.

Disadvantages of the linear and piecewise-linear models are shown in the following. The main objective of this present study is development of a simple real-time transient performance-model for jet engine control. A jet engine arrives to its most dangerous condition during transient operation that may be triggered by fast changes of the input fuel command signal. Thus, control system specifications are formulated for transient response of *maximal variance of input* $\dot{\boldsymbol{m}}_f(\boldsymbol{t})$ (from idle to maximum power level). For example, one spool engine control system specifications include: static error, over-shut, settling time for maximal rotational speed step-function response and constraints for exit turbine temperature, stall margin and equivalence ratio. Basically, small $\dot{m}_f$ variances *do not lead to compressor stall* as the dynamic operating line is located within a short distance from the static (steady state) operating line and do not approach the surge line. There is maximal risk of the compressor stalling as well as of achieving critical turbine temperature and equivalence ratio in transient operation that is triggered by fast variances of $\dot{\boldsymbol{m}}_f(\boldsymbol{t})$. Thus, the linear engine model, describing dynamic processes around an operating point does not provide adequate engine behavior (and plant description) to design a controller for the entire flight envelope. The piecewise-linear models provide greater accuracy for fast simulations over a wide range. However, these are not convenient for controller design due to the following reasons:

Contributed by the IGTI Control, Diagnostics and Instrumentation Division of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received August 29, 2003; final manuscript received May 3, 2004. Assoc. Editor: A. Volponi.

(a) The piecewise-linear model coefficients depend on input/output variables so it is a nonlinear model. This model is nonlinear even at any intermediate stage because "linear" function $y=kx+b$ (k and b are constants) is not linear according to linear system definition (the superposition theorem is not valid). So application of linear control theory may lead to a mistaken design of the *linear* controller throughout the engine operating envelop.

(b) Stability examination of a piecewise-linear control system using linear theory may lead to an incorrect solution. A nonlinear control system may be unstable while the piecewise-linearized control system is stable at all linear "stages" (according to the absolute stability criterion of V. M. Popov).

(c) Sometimes the piecewise-linear model is replaced with a linear plant including uncertainties (linear equations with coefficients slowly changed in any range). Robust control methods are applied for these plants (as Quantitative Feedback Theory). It is a sufficiently complicated and inconvenient plant description of the engine controller design to cause coefficient changes during fast transient operation in the engine model.

(d) Partial derivative calculation for tens of nonlinear functions of the aerothermal model around tens/hundreds of operating points is complicated.

An alternative simple/fast quasilinear engine model aimed for controller design is discussed in this paper. The novel engine model provides maximal accuracy for the largest variance (from minimum to maximum) of the input $\dot{m}_f(t)$ for different M and H. This condition matches the engine control system specifications which are formulated for maximal variance of the model input $\dot{m}_f(t)$. The novel model contains quasilinear differential equations (or transfer functions) and nonlinear algebraic equations.

The fast engine model, discussed in this paper, is developed using the Novel Generalized Describing Function (NGDF) based on Generalized Describing Function (GDF) proposed by the first author for investigation of nonlinear control systems [13]. In [13], the input signal is determined using control system specifications by inverse solution of engine equations. In the present paper the input signal is determined by closed-loop engine simulations. The quasilinear approach of nonlinear functions is used in the form of transfer functions. A simple engine controller was specially designed for these simulations.

The paper presents the GDF and NGDF description and then discusses the application of this technique for the development of fast turbine engine simulation suitable for real-time applications and control. Simulation results between the conventional and fast models are compared and found to provide good agreement.

2 Dynamic Engine Model for Control and Real-Time Simulations

2.1 Conventional Engine Model. Engine aerothermal models based on the thermodynamics of propulsion and mechanical laws contain sets of algebraic and differential equations [1–8]:

$$F_D\left[\frac{dx}{dt},x,y,u,A(x,y,u)\right]=0 \tag{2.1}$$

$$F_A[x,y,u,B(x,y,u)]=0. \tag{2.2}$$

The following variables were used in the "Microjet" model [7]:

$$x=N_{\mathrm{cor}}$$

$$y=\left[N_{\mathrm{cor}},\dot{m}_a,\frac{P_{03}}{P_{02}},T_{05},SM,\phi,\ldots\right]$$

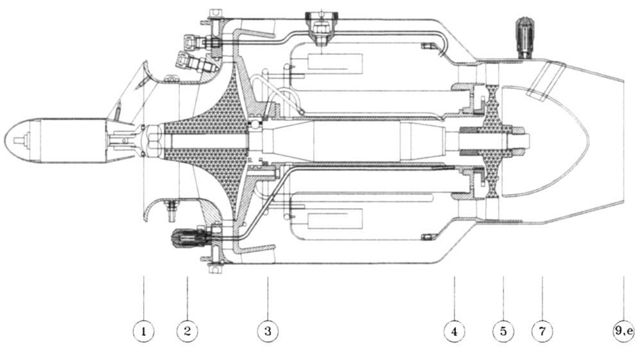


Fig. 1 Engine stations (the drawing refers to the AMT Netherlands B.V. Olympus Design)

$$u=[\dot{m}_f,H,M]. \tag{2.3}$$

Functions A and B of Eqs. (2.1) and (2.2) depend on input/output variables. For example, the differential equation of the engine power balance is

$$C_{p,gt}\cdot\dot{m}_g\cdot(T_{04}-T_{05})\cdot\eta_m-C_{p,ac}\cdot\dot{m}_a\cdot(T_{03}-T_{02})-\mathrm{altpower}=\frac{dE}{dt}, \tag{2.4}$$

where

$$E=\frac{I\cdot\left(\frac{N\cdot\pi}{30}\right)^2}{2}. \tag{2.5}$$

Equation (2.4) is dependent on various thermodynamic properties (i.e., temperatures, air/gas mass flows, etc.). These thermodynamic properties are calculated from the set of algebraic equations. Hence a separate solution of the differential and algebraic equations is impossible. In order to shorten computation time, it is desirable to replace the engine model [Eqs. (2.1) and (2.2)] by an alternative description containing a set of differential equations that are independent of the algebraic equations.

2.2 Fast Engine Model. The following mathematical model of an engine is proposed for a fast numerical solution:

$$\begin{gathered}F_{Dq}\left[\frac{d(\Delta x)}{dt},\Delta x,\Delta u,A_{Dq}(\Delta u)\right]=0\\ F_{Aq}[x,y,u,B_{Aq}(x,y,u)]=0\\ x=x_0+\Delta x\\ u=u_0+\Delta u\end{gathered} \tag{2.6}$$

F_{Dq} = set of differential equations

F_{Aq} = set of algebraic equations

u_0, Δu = vectors of initial values and variances of input variables u

x_0, Δx = vectors of initial values and variances of "state" variables x

y = vector of output variables

A and B = set of functions or constants.

Functions A of the differential equations do not depend on variables x and y of the algebraic equations.

Example of the Fast Model for "Microjet" Engine. The "Microjet" is a "small" one spool jet engine without afterburner with convergent nozzle having maximal thrust of 170 N. For convenience, the station numbers are defined in Fig. 1.

$$u=[\dot{m}_f,H,M]$$

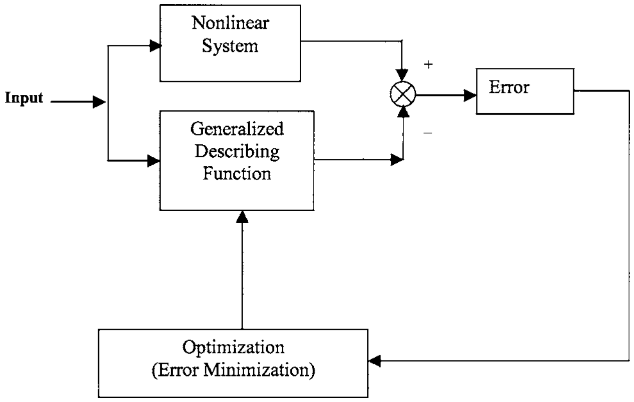


Fig. 2 Generalized describing function definition

$$\Delta x = \left[\Delta N_{\mathrm{corr}}, \Delta \dot{m}_a, \Delta\left(\frac{P_{03}}{P_{02}}\right), \Delta T_{05}, \Delta F \right]$$

$$x_0 = \left[N_{\mathrm{corr}_0}, \dot{m}_{a_0}, \left(\frac{P_{03}}{P_{02}}\right)_0, T_{05_0}, F_0 \right]$$

$$y = \left[N_{\mathrm{corr}}, \dot{m}_{a\mathrm{corr}}, \frac{P_{03}}{P_{02}}, T_{05}, F, SM, \phi, \ldots \right];$$

$$\dot{m}_f = \dot{m}_{f_0} + \Delta \dot{m}_f$$

$$N_{\mathrm{cor}} = N_{\mathrm{cor}_0} + \Delta N_{\mathrm{cor}}$$

$$P_{32} = P_{32_0} + \Delta P_{32}$$

$$F = F_0 + \Delta F$$

$$\dot{m}_a = \dot{m}_{a_0} + \Delta \dot{m}_a$$

$$T_{05} = T_{05_0} + \Delta T_{05} \tag{2.7}$$

The differential equations are:

$$\tau_{N\mathrm{cor}}(H,M,\Delta \dot{m}_f)\frac{d(\Delta N_{\mathrm{cor}})}{dt} + \Delta N_{\mathrm{cor}} = K_{N\mathrm{cor}}(H,M,\Delta \dot{m}_f)\Delta \dot{m}_f$$

$$\tau_{P32}(H,M,\Delta \dot{m}_f)\frac{d(\Delta P_{32})}{dt} + \Delta P_{32} = K_{P32}(H,M,\Delta \dot{m}_f)\Delta \dot{m}_f$$

$$\tau_F(H,M,\Delta \dot{m}_f)\frac{d(\Delta F)}{dt} + \Delta F = K_F(H,M,\Delta \dot{m}_f)\Delta \dot{m}_f$$

$$\tau_{2,\dot{m}_a}(H,M,\Delta \dot{m}_f)\frac{d(\Delta \dot{m}_a)}{dt} + \Delta \dot{m}_a = K_{\dot{m}_a}(H,M,\Delta \dot{m}_f)\left(\tau_{1,\dot{m}_a}(H,M,\Delta \dot{m}_f)\frac{d(\Delta \dot{m}_f)}{dt} + \Delta \dot{m}_f \right)$$

$$\tau_{2,T_{05}}(H,M,\Delta \dot{m}_f)\frac{d(\Delta T_{05})}{dt} + \Delta T_{05} = K_{T_{05}}(H,M,\Delta \dot{m}_f) \times \left(\tau_{1,T_{05}}(H,M,\Delta \dot{m}_f)\frac{d(\Delta \dot{m}_f)}{dt} + \Delta \dot{m}_f \right) \tag{2.8}$$

The algebraic equations are:

$$\left(\frac{P_{02}}{P_a}\right) = \left[1 + \eta_d \frac{\gamma_a - 1}{2} M^2 \right]^{\gamma_a/\gamma_a^{-1}}$$

$$\left(\frac{T_{02}}{T_a}\right) = 1 + \frac{\gamma_a - 1}{2} M^2$$

$$\mathrm{SM} = \frac{\left.\dfrac{(P_{32})_{\mathrm{stall}}}{(\dot{m}_{2a_{\mathrm{cor}}})_{\mathrm{stall}}}\right|_{N_{\mathrm{cor}}=N_{\mathrm{cor}_i}} - \left.\dfrac{P_{32}}{\dot{m}_{2a_{\mathrm{cor}}}}\right|_{N_{\mathrm{cor}}=N_{\mathrm{cor}_i}}}{\left.\dfrac{P_{32}}{\dot{m}_{2a_{\mathrm{cor}}}}\right|_{N_{\mathrm{cor}}=N_{\mathrm{cor}_i}}} \quad \left(P_{32} = \frac{P_{03}}{P_{02}} \right)$$

$$\phi = \frac{\dfrac{\dot{m}_f}{\dot{m}_a}}{\left.\dfrac{\dot{m}_f}{\dot{m}_a}\right|_{\mathrm{stoichiometric}}}$$

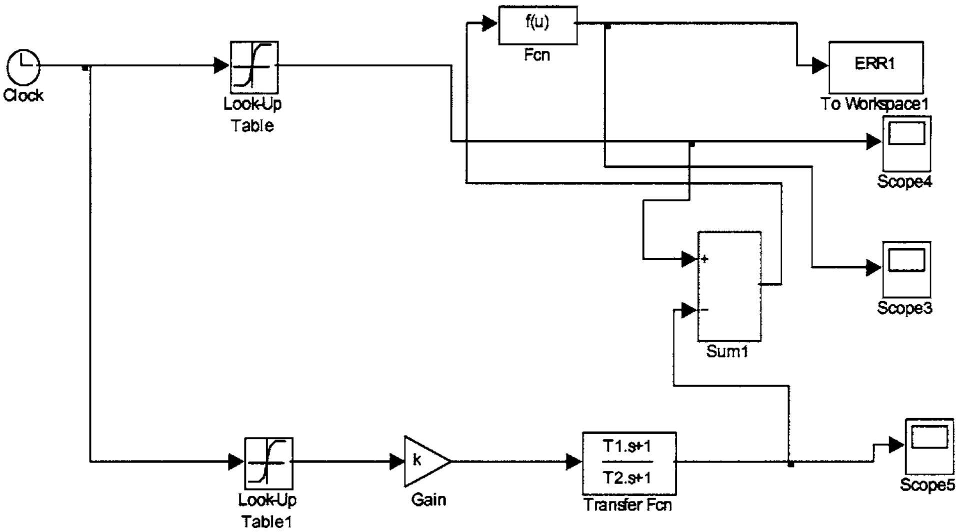


Fig. 3 SIMULINK block diagram for generalized describing function calculations

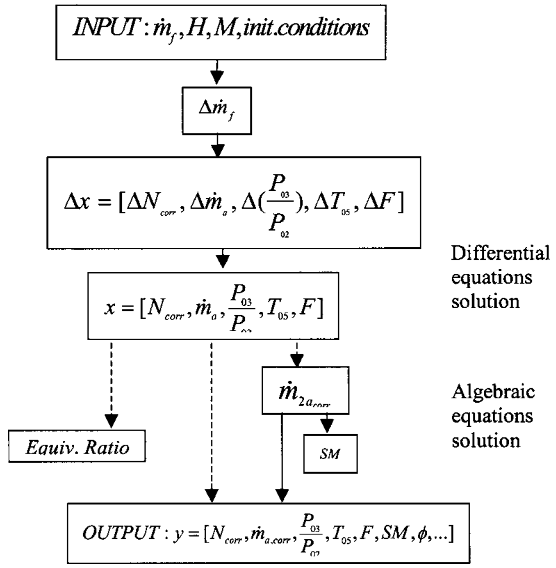


Fig. 4 Real-time model flow chart

$$\dot m_{2a_{\text{cor}}} = \frac{\dot m_{2a}\cdot(1+\varepsilon)\cdot\sqrt{\dfrac{T_{02}}{T_{\text{ref}}}}}{\dfrac{P_{02}}{P_{\text{ref}}}} \qquad (2.9)$$

[In (2.9) we use an SM definition of [4]. One can find another SM definition in [2], p. 472.] The differential Eqs. (2.8) do not depend on variables of the algebraic Eqs. (2.9) so they can be solved separately. The problem is calculation of $K(\Delta\dot m_f, H, M)$ and $\tau(\Delta\dot m_f, H, M)$ in differential Eqs. (2.8). We do not use the linear or piecewise-linear models in this paper. Using the NGDF based on the GDF, is discussed in the following section.

2.3 Novel Generalized Describing Function

2.3.1 Generalized Describing Function. The conventional describing function is used for investigation of periodic oscillations in a nonlinear closed-loop control system. This method is applied for many types of nonlinearities limited by Dirichlet conditions for Fourier series approximation. The other limitation is the "filter hypothesis," i.e., it is assumed that the linear part of the nonlinear control system rejects the highest harmonics of the Fourier series approximation. The transient response of a practical control system, such as an engine, often exhibits damped oscillations before reaching steady state, or may not exhibit any oscillations at all. Using the describing function for investigation of these systems is either very difficult or impossible. The GDF method aimed at investigation of nonlinear closed-loop control systems with an arbitrary input signal is briefly discussed in the present section [13]. It is assumed that the input of an open-loop nonlinear system (such as the jet engine) is known for the GDF calculation. This input may be obtained from control system specifications or from simulations of the closed-loop control system. Accordingly the input signal of a nonlinear open-loop system is approximated for $t\in[t_1,t_2]$ by the following generalized quasipolynomial expression:

$$u(t) = \sum_{i=1}^{k} U_i(a_{0i} + a_{1i}t + a_{2i}t^2 + \cdots + a_{ni}t^n)e^{-\alpha_i t}\cos(\omega_i t + \Theta_i), \qquad (2.10)$$

where U_i, a_{ji}, a_i, ω_i, and Θ_i are constants.

The output $y(t)$ is approximated in this time interval by

$$y'(t) = \sum_{i=1}^{k} U_i e^{-\alpha_i t}[(Q_{0i} + Q_{1a}t + Q_{2i}t^2 + \cdots + Q_{ni}t^n)\cos(\omega_i t + \Theta_i) + (Q'_{0i} + Q'_{1i}t + Q'_{2i}t^2 + \cdots + Q'_{ni}t^n)\sin(\omega_i t + \Theta_i)], \qquad (2.11)$$

where Q_{ji}, Q'_{ji} are constants.

The constant coefficients Q_{ji} and Q'_{ji} are calculated by minimization of errors:

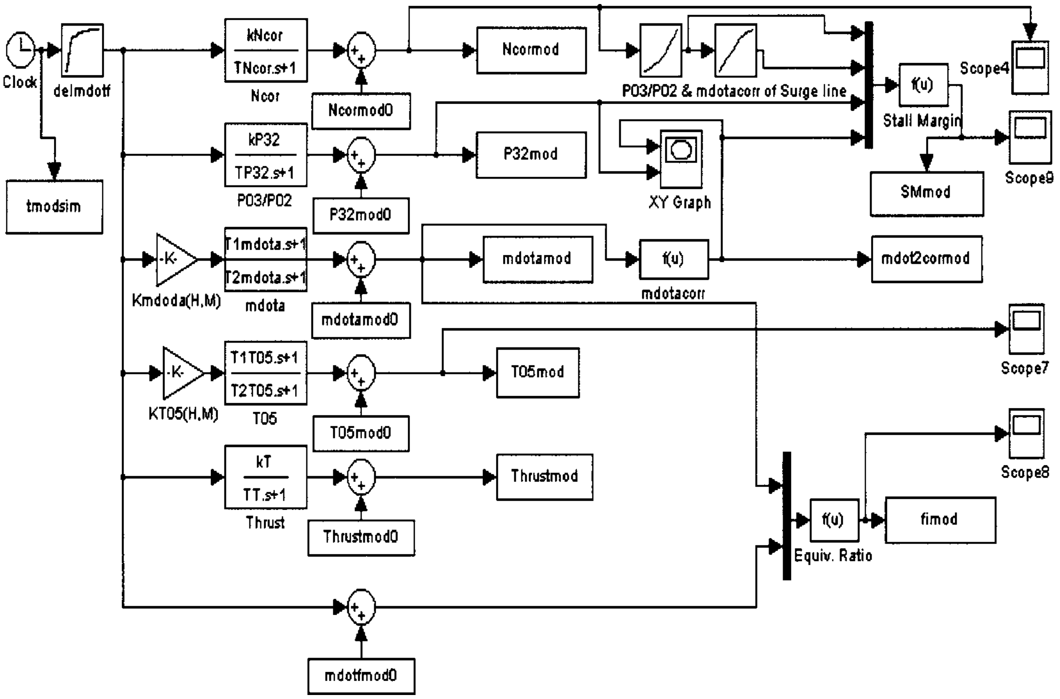


Fig. 5 SIMULINK block diagram for real-time simulation of the open-loop engine

$$\delta^2 = \int_{t_1}^{t_2} [y(t) - y'(t)]^2 dt \tag{2.12}$$

or

$$\Delta = \max|y(t) - y'(t)|. \tag{2.13}$$

The nonlinear system output $y(t)$ is replaced by $y'(t)$.

The GDF is determined as:

$$W(U_i, Q_{0i}, Q_{1i}, \ldots, Q_{ni}, Q'_{0i}, Q'_{1i}, \ldots, Q'_{ni}, \alpha_i, \omega_i, \Theta_i, s) = \frac{Y'(s)}{U(s)}, \tag{2.14}$$

where $Y'(s)$ is the Laplace transform for $Y'(t)$, $U(s)$ is the Laplace transform for input $u(t)$, s is a complex variable.

The nonlinear system is replaced by the "transfer function" (2.14), which depends on the input $u(t)$. The replacement has a small error when the real input signal is similar to the one described in Eq. (2.10).

A simple case is described in the following [n=0 in Eq. (2.11)]:

$$u(t) = Ue^{-\alpha t}\cos(\omega t + \Theta) \tag{2.15}$$

$$y'(t) = Ue^{-\alpha t}[Q\cos(\omega t + \Theta_i) + Q\sin(\omega t + \Theta_i)$$

$$W(U, \alpha, \omega, \Theta, s) = Q(U, \alpha, \omega, \Theta) - Q'(U, \alpha, \omega, \Theta)\frac{s + \alpha}{\omega}. \tag{2.16}$$

Note, for the conventional describing function $\alpha=0$ thus

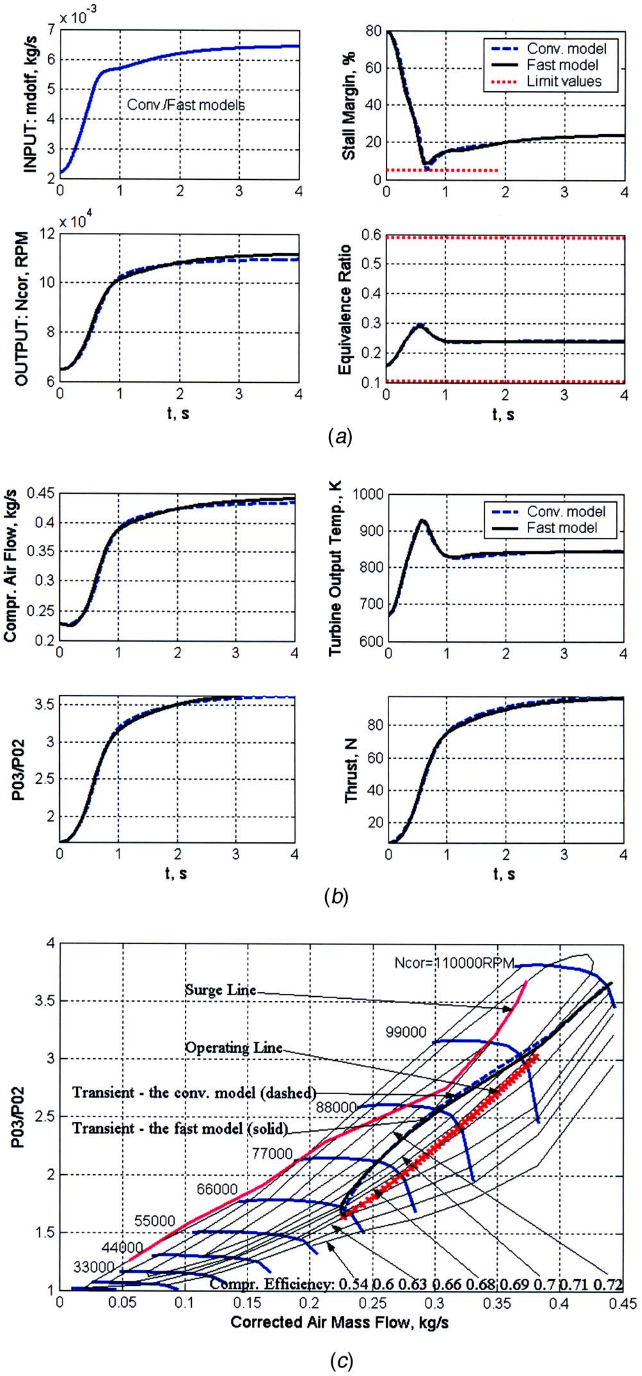
$$u = U\cos(\omega t) \tag{2.17}$$

$$W(U, \alpha, \omega, \Theta, s) = Q(U, \omega) - Q'(U, \omega)\frac{s}{\omega}. \tag{2.18}$$

The GDF method is effective for many types of nonlinearities limited by the minimization of error (2.12) (or (2.13)) and by "generalized filter hypothesis" i.e. filtration of y' in (2.11) and (2.15) at different time intervals [13].

All real "plant" transfer functions have the complex variable s in the denominator (in the characteristic polynomial). The GDFs (2.14) and (2.16) do not have the complex variable s in the denominator. This disadvantage leads to some problem in controller design. In addition, approximation of the open-loop system input *u(t)* in the form of (2.10) [or (2.15)] is complicated.

2.3.2 Application of the Novel Generalized Describing Function for Jet Engine Fast Model. The NGDF, in the form of a "conventional" transfer function (having a characteristic polynomial with complex variable *s*), is suggested in the present section. The approximations (2.10) and (2.11) of the nonlinear system input/output are not used in this method. *NGDF is a quasilinear dynamic model of a nonlinear system (or non-linear function) for an arbitrary input signal*. As for conventional describing function and for GDF, parameters of NGDF depend on the open loop system input. We assume that the input (fuel mass flow) of the open-loop engine is known. This may be obtained from control system specifications using inverse solution of "plant" equations for the known closed loop output signal, or using closed loop simulations (if it is possible). In this paper we use the second method to obtain a "large" input signal. To determine NGDF for the "Microjet," the fuel mass flow vs time was simulated using the conventional nonlinear model of the closed-loop jet engine for maximal variance of closed loop input $N_{\text{corr,ref}}$ (from minimal to maximal values) at different M and H. This conventional engine model is described by nonlinear differential equations of unsteady power balance including rotor moment of inertia (2.4) and algebraic aerothermal equations [7]. A simple lead-lag controller has been previously designed for these closed-loop simulations. Examples of the simulated input $\dot{m}_f$, as well as "state" variables N_{corr}, $\dot{m}_a$, P_{03}/P_{02}, T_{05}, F vs time, are presented in Figs. 6–9 for different M and H (dotted lines for the conventional engine model). All unknown parameters τ and K in Eqs. (2.8) were calculated using the simulated data. Equations (2.8) were rewritten in the form of transfer functions using the Laplace transform:



(*a*)

(*b*)

(*c*)

Fig. 6 (*a*) Comparison of fuel flow (input), corrected rotational speed (output), stall margin, and equivalence ratio of the conventional/fast engine models at *H*=152 m, *M*=0.5. (*b*) Comparison of air flow, compressor pressure ratio, turbine output temperature, and engine thrust for the conventional/fast models at *H*=1524 m, *M*=0.5. (*c*) Comparison of transient operating lines in the compressor map for the conventional/fast engine models at *H*=1524 m, *M*=0.5.

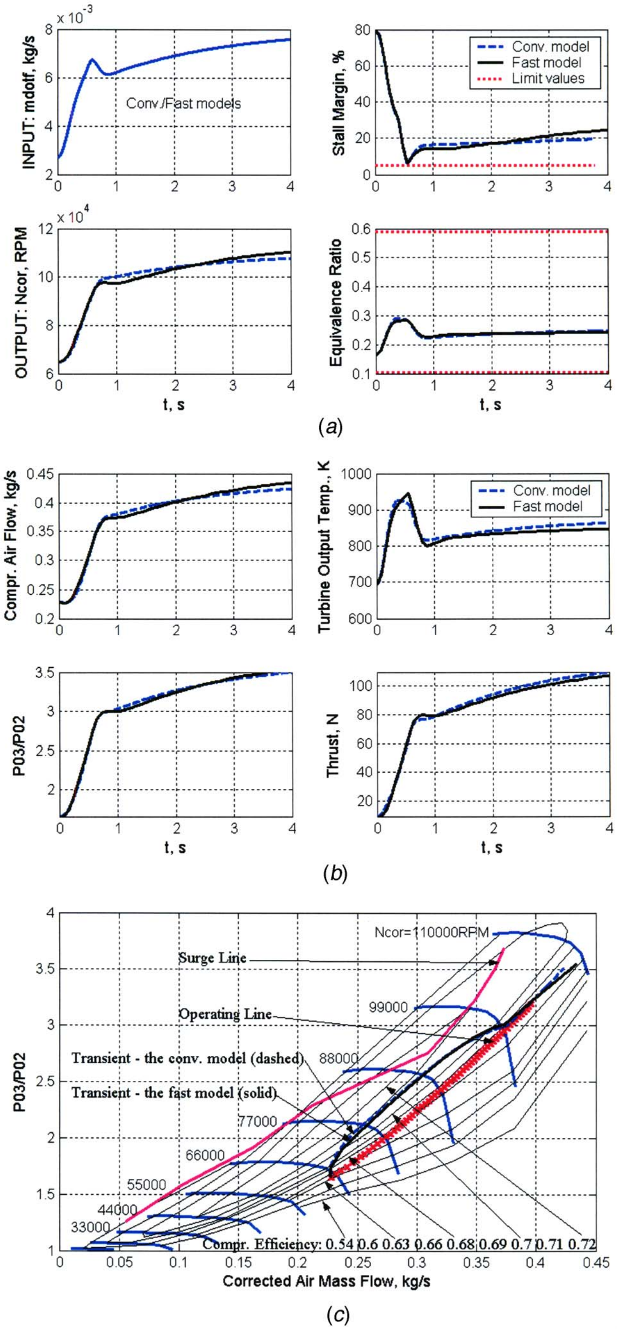


Fig. 7 (*a*) Comparison of fuel flow (input), corrected rotational speed (output), stall margin, and equivalence ratio of the conventional/fast engine models at sea level and *M*=0.5. (*b*) Comparison of air flow, compressor pressure ratio, turbine output temperature, and engine thrust for the conventional/fast models and *M*=0.5. (*c*) Comparison of transient operating lines in the compressor map for the conventional/fast engine models (sea level, *M*=0.5).

$$W_{N_{\mathrm{cor}}}(s,H,M,\Delta\dot{m}_f)=\frac{\Delta N_{\mathrm{cor}}(s)}{\Delta\dot{m}_f(s)}=\frac{K_{N_{\mathrm{cor}}}(H,M,\Delta\dot{m}_f)}{\tau_{N_{\mathrm{cor}}}(H,M)s+1}$$

$$W_{P_{32}}(s,H,M,\Delta\dot{m}_f)=\frac{\Delta P_{32}(s)}{\Delta\dot{m}_f(s)}=\frac{K_{P_{32}}(H,M,\Delta\dot{m}_f)}{\tau_{P_{32}}(H,M,\Delta\dot{m}_f)s+1}$$

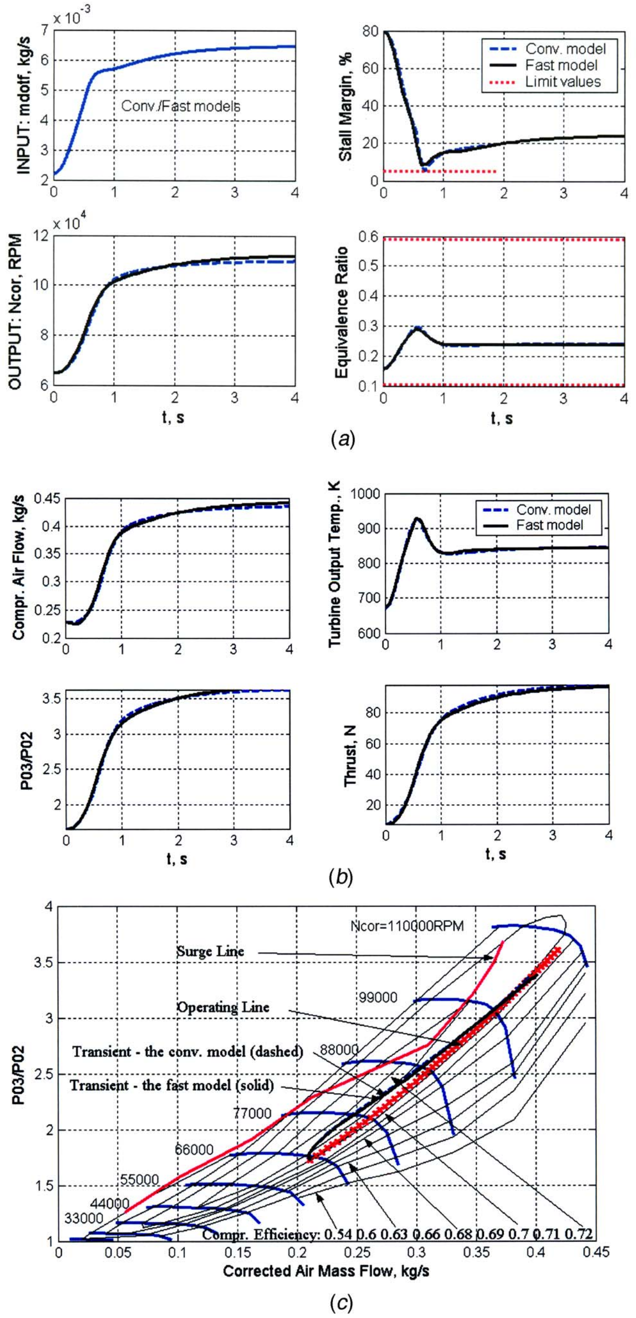
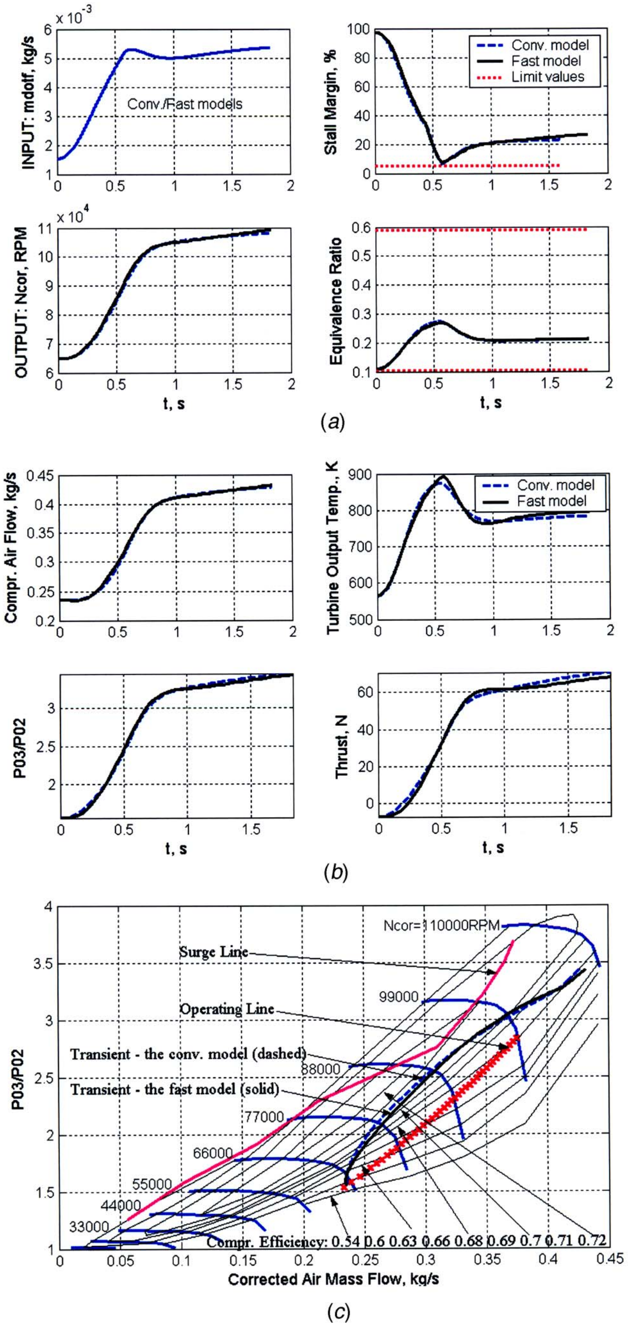


Fig. 8 (*a*) Comparison of fuel flow (input), corrected rotational speed (output), stall margin, and equivalence ratio of the conventional/fast engine models at *H*=912 m, *M*=0. (*b*) Comparison of air flow, compressor pressure ratio, turbine output temperature, and engine thrust for the conventional/fast models at *H*=912 m, *M*=0. (*c*) Comparisons of transient operating lines in the compressor map for the conventional/fast engine models at *H*=912 m and *M*=0.

$$W_{N_{\mathrm{cor}}}(s,H,M,\Delta\dot{m}_f)=\frac{\Delta F}{\Delta\dot{m}_f(s)}=\frac{K_F(H,M,\Delta\dot{m}_f)}{\tau_F(H,M,\Delta\dot{m}_f)s+1}$$

$$W_{\dot{m}_a}(s,H,M,\Delta\dot{m}_f)=\frac{\Delta\dot{m}_a(s)}{\Delta\dot{m}_f(s)}=K_{\dot{m}_a}(H,M,\Delta\dot{m}_f)\times\frac{\tau_{1,\dot{m}_a}(H,M,\Delta\dot{m}_f)s+1}{\tau_{2,\dot{m}_a}(H,M,\Delta\dot{m}_f)s+1}$$



(*a*)

(*b*)

(*c*)

Fig. 9 (*a*) Comparison of fuel flow (input), corrected rotational speed (output), stall margin, and equivalence ratio of the conventional/fast engine models at *H*=3040 m and *M*=0.7. (*b*) Comparison of air flow, compressor pressure ratio, turbine output temperature, and engine thrust for the conventional/fast models at *H*=3040 m and *M*=0.7. (*c*) Comparison of transient operating lines in the compressor map for the conventional/fast engine models at *H*=3040 m and *M*=0.7.

$$W_{T_{05}}(s,H,M,\Delta\dot{m}_f)=\frac{\Delta T_{05}(s)}{\Delta\dot{m}_f(s)}=K_{T_{05}}(H,M,\Delta\dot{m}_f)\times\frac{\tau_{1,T_{05}}(H,M,\Delta\dot{m}_f)s+1}{\tau_{2,T_{05}}(H,M,\Delta\dot{m}_f)s+1}. \tag{2.19}$$

Each W term in Eqs. (2.19) is quasilinear NGDF and dependent on M, H, $\Delta\dot{m}_f$. Unknown parameters τ and K of the NGDFs in Eq. (2.19) are calculated for maximal input $\Delta\dot{m}_f$ at different M and H using an error minimization algorithm presented in Fig. 2. Error (2.13) is preferable to error (2.12) as it reduces the maximal error to the smallest possible value while the minimization of the error (2.12) leads to greater value of the maximal error. NGDF, as GDF, is capable of being used for many types of nonlinearities limited by the minimization of errors (2.12) or (2.13). The condition of "generalized filter" for the linear part of a nonlinear system is not required for NGDF application.

A SIMULINK block diagram for calculation of parameters τ_1, τ_2 and K for the "lead-lag" NGDF

$$W(s)=K(H,M)\frac{\tau_1(H,M)s+1}{\tau_2(H,M)s+1} \tag{2.20}$$

is shown in Fig. 3. The following $\tau(H,M)$ and $K(H,M)$ for (2.19) were calculated using closed loop simulations, presented in Figs. 6–9 by dashed lines:

$$\begin{aligned}K_{N_{\mathrm{cor}}}(H,M)=[&1.20\ 1.10\ 0.94\ 0.75\\ &1.47\ 1.34\ 1.10\ 0.94\\ &1.80\ 1.57\ 1.30\ 1.18\\ &3.22\ 2.69\ 2.12\ 1.80]^{*}1e7;\quad [\mathrm{rpm}/(\mathrm{kg/s})]\end{aligned} \tag{2.21}$$

$$\begin{aligned}\tau_{N_{\mathrm{cor}}}(H,M)=[&0.230\ 0.208\ 0.175\ 0.147\\ &0.235\ 0.209\ 0.170\ 0.155\\ &0.291\ 0.234\ 0.195\ 0.206\\ &0.524\ 0.384\ 0.302\ 0.328];\quad (\mathrm{s})\end{aligned} \tag{2.22}$$

$$\begin{aligned}K_F(H,M)=[&2.81\ 2.38\ 2.02\ 1.79\\ &2.90\ 2.40\ 2.09\ 1.95\\ &2.91\ 2.53\ 2.15\ 1.98\\ &3.34\ 2.72\ 2.20\ 2.07]^{*}1e4;\quad [\mathrm{N}/(\mathrm{kg/s})]\end{aligned} \tag{2.23}$$

$$\begin{aligned}\tau_F(H,M)=[&0.317\ 0.303\ 0.172\ 0.110\\ &0.390\ 0.311\ 0.200\ 0.121\\ &0.390\ 0.361\ 0.220\ 0.159\\ &0.760\ 0.573\ 0.266\ 0.190];\quad (\mathrm{s})\end{aligned} \tag{2.24}$$

$$\begin{aligned}K_{P_{32}}(H,M)=[&4.73\ 4.48\ 3.92\ 3.34\\ &5.95\ 5.47\ 4.74\ 4.08\\ &7.53\ 6.73\ 5.76\ 4.99\\ &12.30\ 10.90\ 9.11\ 7.94]^{*}100;\quad [1/(\mathrm{kg/s})]\end{aligned} \tag{2.25}$$

$$\begin{aligned}\tau_{P_{32}}(H,M)=[&0.286\ 0.286\ 0.191\ 0.153\\ &0.364\ 0.300\ 0.218\ 0.160\\ &0.463\ 0.333\ 0.239\ 0.193\\ &0.676\ 0.501\ 0.356\ 0.298];\quad (\mathrm{s})\end{aligned} \tag{2.26}$$

$$\begin{aligned}K_{\dot{m}_a}(H,M)=[&52.35\ 49.20\ 45.46\ 43.42\\ &52.94\ 49.90\ 45.03\ 43.35\\ &55.66\ 49.23\ 44.90\ 45.49\\ &58.47\ 53.56\ 47.00\ 45.83];\end{aligned} \tag{2.27}$$

$$\begin{aligned}\tau_{1,\dot{m}_a}(H,M) = -[&0.0721\ 0.0650\ 0.0668\ 0.0369\\ &0.1940\ 0.1700\ 0.1150\ 0.0769\\ &0.2690\ 0.1940\ 0.1242\ 0.0730\\ &0.3740\ 0.3060\ 0.2090\ 0.1440]; \qquad (2.28)\end{aligned}$$

$$\begin{aligned}\tau_{2,\dot{m}_a}(H,M) = [&0.2700\ 0.2740\ 0.2060\ 0.1700\\ &0.2250\ 0.2060\ 0.1660\ 0.1660\\ &0.2210\ 0.2030\ 0.1900\ 0.2250\\ &0.2440\ 0.2480\ 0.2060\ 0.2590]; \qquad (2.29)\end{aligned}$$

$$\begin{aligned}K_{T_{05}}(H,M) = [&1.41\ 2.50\ 3.03\ 3.50\\ &1.56\ 2.68\ 4.03\ 4.85\\ &2.10\ 3.46\ 4.94\ 5.84\\ &4.16\ 5.73\ 7.19\ 8.37]^{*}1e4; \quad [\mathrm{K/(kg/s)}] \qquad (2.30)\end{aligned}$$

$$\begin{aligned}\tau_{1,T_{05}}(H,M) = [&2.88\ 1.56\ 0.893\ 0.460\\ &3.74\ 1.92\ 0.682\ 0.372\\ &3.82\ 1.63\ 0.753\ 0.415\\ &2.93\ 1.76\ 1.040\ 0.560]; \quad (\mathrm{s}) \qquad (2.31)\end{aligned}$$

$$\begin{aligned}\tau_{2,T_{05}}(H,M) = [&0.186\ 0.189\ 0.210\ 0.154\\ &0.163\ 0.150\ 0.128\ 0.113\\ &0.134\ 0.165\ 0.172\ 0.125\\ &0.105\ 0.156\ 0.226\ 0.187] \quad (\mathrm{s}) \qquad (2.32)\end{aligned}$$

In these matrices, rows match to:

$$\frac{H}{H_{\max}}0;\ 0.25;\ 0.50;\ 1.00, \qquad (2.33)$$

columns match to:

$$\frac{M}{M_{\max}}0;\ 0.3;\ 0.6;\ 1.0. \qquad (2.34)$$

The maximal instantaneous relative error of these calculations, as determined in Figs. 2 and 3, is 0.03.

We used *maximal variance* of closed loop input (from idle to maximal power level) for the engine simulations presented in Figs. 6–9 to provide the maximal accuracy of the simple engine model and to design the controller for this "dangerous" state. Parameters of the controller and of the fast model are calculated in the following iterative process. In each iteration step the new controller is generated using the fast model received from closed-loop simulations using the controller generated in the previous iteration step. In practice, two to three iteration steps are required for the model and controller design. Finally the controller parameters are defined more precisely using closed loop simulations. (This process of controller design has not been described because of size limitations of the present paper and will be these subject of subsequent publication.)

What happens when the input signal dependent upon the scenario, pilot inputs, etc. is not similar to the maximal signal used for the fast model design? If magnitude of the current input signal is distinguished from this maximal signal, during the engine operation, then the model error increases. The error does not increase significantly when the input signals differ slightly—from the maximal so time derivatives of the nonlinear engine model functions does not have discontinuity points (these nonlinearities are called "weak"). Whenever the engine fuel input command magnitude is significantly less than the maximal, the engine operational state is stable and there is less danger for stall. Therefore less model accuracy is required for the controller design. To decrease this error parameters τ and K in Eq. (2.19) are calculated using different input variances $\Delta\dot{m}_f$. This process is like the conventional describing function or GDF calculation when parameters depend on the input magnitude (it is also like the piecewise-linear model calculation). In this case, matrices (2.21)–(2.32) are three-dimensional. These more complicated models are "linked" together via an interpolation scheme based upon scheduling parameters $\dot{m}_f$, M, and H.

3 Engine Real-Time Simulations

A flow chart of the program for one spool engine real-time calculations using the fast model is shown in Fig. 4. One can see that the differential Eq. (2.8) and algebraic Eq. (2.9) equations are calculated separately. A SIMULINK block diagram for the simulation is presented in Fig. 5. Each differential equation of (2.8) is coded in the transfer function form. Examples of open-loop real-time simulation are presented in Figs. 6–9 by the solid lines. In these simulations input ($\dot{m}_f$) was matched to the engine transient operation from idle (N_{cor}=65,000 rpm) to maximal power level (N_{cor}=110,000 rpm). Computation time was found to require 1–2 s for the 2–4 s engine process (Pentium III). Hence we performed *real-time simulation*. Comparison between simulations using the fast and conventional engine models is shown in Figs. 6–9. We used the conventional model of open loop one-spool jet engine described in [7] and Matlab/Simulink program [8] for these simulations. The conventional engine model includes the differential equation (2.4) and 58 nonlinear aerothermal algebraic equations. The differential equation integration was performed in SIMULINK using the Euler integration scheme with constant integration step of 2 ms. Nonlinear algebraic equations are solved several times during the numerical solution of the differential equation. An iterative balance process is used to ensure convergence of the algebraic equations solution within specific tolerances (0.0005) for each integration step. The conventional model was run on the same computer as the fast engine model. The conventional model computation time was 200–400 s. One can see in Figs. 6–9 that the fast engine model relative error is less than 3%–5% for all variables.

The novel jet engine quasilinear model is no more complicated than the linear engine model. It describes, with utmost precision, the dynamic processes "around" the large inputs (from idle to maximum power level) while the linear engine model has the most precision for small variances around an operating point. Thus, the novel engine model is preferable for jet engine controller design.

4 Conclusions

(1) The NGDF method is capable of creating a simple quasi-linear engine model using large input variances, being different to the linear engine model using small input variances around an operation point. This novel model is meant to simplify controller design according to the control system specifications which declare the limiting parameters of maximal input response in transient operation. The fast model can be used for real-time engine simulation. However, it may become relatively complicated for precise real-time engine simulations with full-envelope coverage for all possible inputs. However it is not more complicated than the piecewise-linear model.

(2) Computation time of the fast engine model, coded in SIMULINK, was found to require 1–2 s for the 2–4 s of engine process while the computation time for the conventional engine model was 200–400 s for the same process simulation using the same implementation environments. Thus, the real-time model requires less computation time than the

physical process in a real engine by a factor of 2 and requires less computation time than the conventional engine model by a factor of hundreds.

(3) Comparisons between conventional and fast model simulations at flight envelope: $0 < H < H_{\max}$ and $0 < M < M_{\max}$ prove the validity of the fast model for large input variances.

Acknowledgment

We would like to thank the Israel Ministry of Immigration and Absorption for financial support and the reviewers for their invaluable advice.

Nomenclature

A = set of functions or constants in nonlinear engine model
a_{ji} = constant coefficients in quasipolynomial approximation of nonlinear system input
altpower = alternator power
B = set of functions or constants in nonlinear engine model
C_p = specific heat at constant pressure
E = kinetic energy of engine rotor, (J)
F = thrust (N)
F_A = set of algebraic equations in the conventional engine model
F_D = set of differential equations in the conventional engine model
F_{Aq} = set of algebraic equations in the fast engine model
F_{Dq} = set of differential equations in the fast engine model
GDF = generalized describing function
f = fuel/air ratio
H = altitude, (feet) or (m)
I = engine rotor moment of inertia (kg m^2)
K = proportional component of transfer function
M = Mach number
$\dot{m}_a$ = air mass flow at station 3 (compressor exhaust) (kg/s)
N = rotational speed (rpm)
NGDF = Novel generalized describing function
P_{oi} = total pressure at station i (Pa)
SM = stall margin
s = complex variable
t = time (s)
T_{0i} = total temperature at station i (K)
U = "magnitude" of exponent-harmonic signal
$U(s)$ = Laplace transform for input $u(t)$
u = input of nonlinear system
Q, Q' = generalized describing function coefficients
W = transfer function, generalized describing function
x = vector of "state" (intermediate) variables
$Y'(s)$ = the Laplace transform for y'
y = vector of nonlinear system output variables
y' = quasipolynomial approximation of nonlinear system output

Greek

α = damping coefficient (1/s)
δ^2 = square error
γ_a = specific heat ratio
Δ = "large" variance or absolute error
η = efficiency
ε = air leakage, fraction of compressor airflow leakage and inlet airflow
ϕ = equivalence ratio
τ = time constant (s)
ω = angular velocity (1/s)
Θ = phase of harmonic signal (rad)

Subscripts

0 = initial value
A = algebraic equation of the engine model
Aq = algebraic equation of the engine fast model
a = air
c = compressor
cor = corrected parameter
D = differential equation of the engine model
Dq = differential equation of the engine fast model
f = fuel
g = gas
i = station number
n = nozzle and quasipolynomial power
m = mechanical (efficiency)
ref = reference value
t = turbine

References

[1] MacIsaak, B. D. and Saravanamutto, H. I. H., 1974, "A Comparison of Analog, Digital and Hybrid Computing Techniques for Simulations of Gas Turbine Performance," ASME Paper No. 74-GT-127.
[2] Walsh, P. P., and Fletcher, P., 1998, *Gas Turbine Performance*, Blackwell Science, USA.
[3] Lichtsinder, M. and Levy, Y., 2002, "Two-Spool. Turbo-Fan Engine. (A) Steady-State and Dynamic Mathematical Models," TAE No. 903, Technion, Israel, pp. 20–37.
[4] Khalid, S. J. and Hearne, R. E., 1980, "Enhancing Dynamic Model Fidelity for Improved Predication of Turbofan Engine Transient Performance," AIAA 80-1083.
[5] Sellers, J. F. and Daniele, C. J., 1975, "DYNGEN — A Program for Calculating Steady-State and Transient Performance of Turbojet and Turbofan Engines," Lewis Research Center, National Aeronautics and Space Administration, pp. 2–17.
[6] Jones, A. C., 1983, "NPT-171 Turbojet Performance Prediction Method," Noel Penny Turbine Ltd., internal publication, pp. 2–21.
[7] Lichtsinder, M. and Levy, Y., 2002, "Steady-State and Transient Performance of Single-Spool Turbojets Using Novel MATLAB Program," Int. J. Turbo Jet Engines, **19**(3), 139–156.
[8] Lichtsinder, M. and Levy, Y., 2001, "Matlab Prediction Program for Steady-State and Transient Performance of Single-Spool Turbojets," 41st Israel Annual Conference on Aerospace Sciences, Israel.
[9] French, M. W., 1982, "Development of Compact Real-Time Turbofan Engine Dynamic Simulation," SAE 821401.
[10] Sugiyama, N., 1990, "Generalized High Speed Simulation of Gas Turbine Engines," ASME 90-GT-270.
[11] Vittal Rao, S., Moellenhoff, D., and Jaeger, J. A., 1990, "Linear State Variable Dynamic Model and Estimator Design for Allison T406 Gas Turbine," Int. J. Turbo Jet Engines, **7**, pp. 179–186; also in Conf. Proceedings of 1988 Gas Turbine and Aeroengine Congress.
[12] Prakash, Rajiva, Vittal Rao, S., and Skarvan, C. A., 1990, "Life Enhancement of a Gas Turbine Engine by Temperature Control," Int. J. Turbo Jet Engines, **7**, pp.187–205.
[13] Lichtsinder, M., 1982, "Generalized Method for Nonlinear Control System Calculations," Uzbek Academy of Sciences, pp. 3–48 (in Russian).

Fredrik Haglind
FOI, Swedish Defence Research Agency,
Systems Technology,
SE-164 90 Stockholm, Sweden

Riti Singh
Cranfield University,
School of Engineering,
Cranfield, MK 430AL, UK

Design of Aero Gas Turbines Using Hydrogen

Mainly owing to the dwindling fossil oil resources and the environmental concerns of discharging greenhouse gases into the atmosphere, it is essential to find an alternative to kerosene for civil aviation. This paper covers the main effects on aero engines when changing to hydrogen fuel. Particularly, emission and performance issues are discussed, but some design matters are also covered. By simply changing to hydrogen, small engine performance gains may be obtained. The results of the calculations suggest that there is the potential to design a combustion system using hydrogen that produces less NO_x *emissions than any system using kerosene.* [DOI: 10.1115/1.2179468]

1 Introduction

Globally, passenger aviation traffic is expected to grow, averaging around 5.3% annually over the next 20 years [1], implying an almost threefold traffic volume at the end of this period. At the same time, the greenhouse effect generated by emissions produced by human activity, particularly carbon dioxide, has become an increased concern. The majority of scientists today are in agreement that discharging greenhouse gases into the atmosphere has an impact on the global climate. Furthermore, the dwindling fossil oil resources raise concerns. Finding an alternative to kerosene is thus essential.

Taking a completely new approach for civil aviation by fueling with liquid hydrogen, in the longer term produced from renewable energy sources, would enable compatibility with the environment. The civil aviation would be powered by an energy carrier that may be produced from any energy source, by electrolysis of water. It would reduce the unhealthy emissions in the airport vicinity, and, even more importantly, imply a significant reduction of civil aviation's contribution to global warming. When hydrogen is burned the emissions consist only of water vapor (H_2O) and oxides of nitrogen (NO_x). All emissions containing carbon and sulphur are eliminated. However, looking from a systems perspective, the hydrogen usage may, depending on the hydrogen production method, imply that also other pollutant emissions than H_2O and NO_x are generated.

Although offering great prospects, use of liquid hydrogen as an aviation fuel poses formidable challenges regarding technical development, development of feasible, efficient hydrogen production methods, handling, and aircraft design. As for the aero engine, the cycle and combustion chamber design need careful attention to secure a safe and reliable operation, as well as to exploit fully the favorable properties offered by using hydrogen for aero gas turbines. The engine could be designed either with a minimum number of hardware changes or by employing unconventional cycles, which exploit the cryogenic properties of hydrogen, in order to improve the performance.

This paper covers the main effects on aero gas turbines when changing to hydrogen fuel, with emphasis on environmental issues. The objective is to answer the question of whether there are any significant gains, in terms of engine performance and pollutant emissions, as well as to assess the technical feasibility of changing to hydrogen fuel. For reasons discussed below, there might be the potential for lowering the NO_x emissions compared to conventional kerosene combustors. Using simulation tools, the potential of achieving low-NO_x emissions is illustrated for a conventional as well as for a hydrogen-fueled engine by calculating the flame temperature for relevant operating conditions. Different engine cycle layouts are evaluated, and the consequences on design-point performance are quantified. Design and operational matters of particular interest when using hydrogen are also discussed.

This paper is divided into five sections. In Sec. 2, the issues related to engine performance are raised; in Sec. 3, the effects on pollutant emissions of using hydrogen are covered; in Sec. 4, some design and operational matters of interest when using hydrogen are discussed; and finally, the conclusions are summarized in Sec. 5.

2 Engine Performance

When burning hydrogen in an aero gas turbine there are several issues that need to be regarded. In addition to redesigning the combustion chamber, the minimum change that needs to be adopted with the engine cycle is the implementation of facilities to evaporate the hydrogen (which is stored in the tanks in a liquid state) prior to its entry into the combustion chamber. The fuel heating can be accomplished either by an external heat source or a heat exchanger (HE) located at a suitable engine location. Placing the HE outside the engine does not affect the engine performance; however, it might cause practical problems when the fuel system is integrated with the other aircraft systems. Available external heat sources are, for instance, cooling systems of hydraulic fluids, pumps, electric equipment, and the cabin area [2].

Looking at the option of employing a HE, there are various possible engine locations that have been studied within the European Commission–(EC)-sponsored project, CRYOPLANE [2–5]. When deciding where to place the HE, possible benefits in performance need to be weighed against increased engine complexity and safety issues. Aiming at minimizing the number of hardware changes and prioritizing safety issues, a HE located at the LPT (low-pressure turbine) exit is employed here. Other possible engine configurations are discussed in Sec. 2.4.

In Sec. 2.1, the engine used in the simulations is described. The heat exchanger is analyzed in Sec. 2.2, including an approximate assessment of the heat-transferring area that would be required if the struts in the exhaust are to be employed to heat up the hydrogen. A detailed design of a heat exchanger is beyond the scope of this paper. Effects on performance are illustrated in Sec. 2.3 by carrying out engine simulations. In addition, a number of suggested unconventional engine cycle configurations are described and discussed (Sec. 2.4).

2.1 Choice of Engine for Simulations.

In order to investigate the consequences on performance when changing to hydrogen, a particular engine, namely, the V2527-A5 engine, is simulated.

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received October 1, 2003; final manuscript received March 1, 2004. Review conducted by A. J. Strazisar. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Vienna, Austria, Paper No. GT2004-53349.

This is a two-shaft boosted turbofan engine suitable for short- and medium-range aircraft. The jets from the bypass and core are unmixed. It features a single-stage fan, a four-stage low-pressure (LP) compressor and a ten-stage high-pressure (HP) compressor. The HP inlet guide vane and the first three vane stages are variable (not modeled in the calculations). The HP and LP turbines feature two and five stages, respectively. The reason for choosing this engine is that it is a typical, reasonably modern, medium-thrust turbofan engine and that there are sufficient public data on it to create a sufficiently detailed model of its performance. Furthermore, this engine is suitable for the main study aircraft within the Cryoplane project, namely, the A320, on which this work is based.

With small modifications (lower HP-compressor (HPC) isentropic efficiency and inclusion of losses in the exhaust duct), the model is based on the one proposed within the CRYOPLANE project [2,4,5]. Also, as opposed to their model, a thrust coefficient, which is calculated depending on the expansion ratio of the nozzle and hence varies with the operating condition, is applied here. Using this methodology, a thrust coefficient equal to about 0.98 is applied. In total, therefore, a slightly decreased performance in terms of specific fuel consumption (SFC) and specific thrust (SPT) is obtained here compared to that presented in the CRYOPLANE project. The main design parameters at the design point are as follows:

- Design point: takeoff (T/O), static sea level, ISA+10 K
- BPR: 4.8
- OPR: 28.5 (1.5×2.0×9.5)
- Thrust: 117.9 kN

Reasonable component efficiencies are applied and pressure losses in intake, bypass duct, combustor, and exhaust duct are included. Neither overboard bleed nor power extraction is included.

To enable higher turbine entry temperatures than the maximum allowable metal temperature, cooling of hot parts is required. This is accomplished by bleeding off a part of the compressed air which then passes through cooling passages inside the blades. Generally, the NGV (nozzle guide vane) and HP turbine rotor need to be cooled. Also air needs to be taken off for sealing, aiming to stop expanding gases from penetrating the disk system. In order to simulate these effects, the same numbers as applied by Boggia et al. [2] are applied. It is assumed that 15% of the core air is bled off after the HP compressor. Depending on where expansion work takes place, this air is mixed with the hot gases at different points along the expansion. To simulate NGV cooling, 12.5% of the bleed is injected immediately before the HP turbine, and the remaining 2.5% is injected immediately before the LP turbine to be used for HP rotor cooling and seals.

2.2 Evaluation of Heat Exchangers. Practically, the heat exchanging might be accomplished by employing the struts, which are the mechanical structures in the exhaust that hold the rotors in place and are connecting the outer structure of the bearings. These are already-existing structures that do not add any pressure loss and do not cause any significant changes in engine complexity or engine weight. The idea is to pass the hydrogen through the hollow struts, which are heated on the outside by the exhaust gases. There are, however, safety issues associated with this design, due to the potential detrimental consequences of a fuel leak. In order to employ this idea, the design would have to meet the airworthiness requirement. Alternatively, the HE could feature a simple coil tube placed over the inside face of the jet pipe casing, avoiding major engine changes and giving a relatively aerodynamically clean jet pipe [3].

In order to obtain a stable combustion and to avoid the critical temperature where density and viscosity of the hydrogen varies essentially with temperature (which is ~60 K), the hydrogen needs to be heated up prior to its entry into the combustion chamber. Furthermore, hydrogen reliquefaction across the fuel nozzle due to the pressure drop is prevented by the fuel preheating. For these reasons, it was suggested within the CRYOPLANE project that the temperature of the hydrogen prior to the combustion chamber should exceed 150 K [3]. In the subsequent calculations, a fuel injection temperature of 250 K is used.

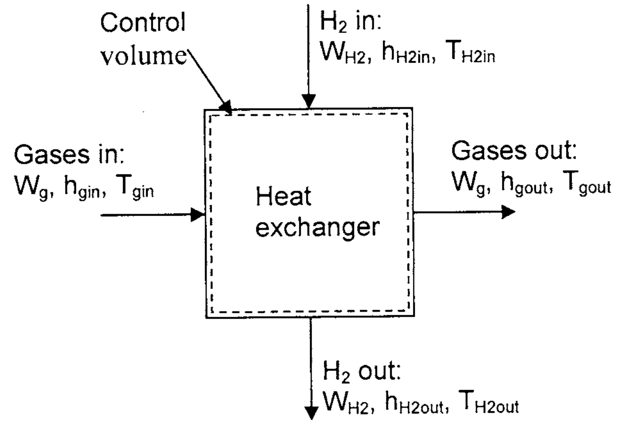


Fig. 1 Notations for a heat exchanger to evaporate the hydrogen

The energy necessary to raise the fuel temperature to the required value is taken into account by decreasing the exhaust gases' total temperature before the nozzle expansion, which, in turn, decreases the thrust output slightly. The temperature drop is calculated by employing the conservation of energy equation for a steady-flow system [6], with the notations given in Fig. 1

$$Q - E = \sum W_{\mathrm{out}} h_{\mathrm{out}} - \sum W_{\mathrm{in}} h_{\mathrm{in}} \tag{1}$$

where Q denotes heat transfer, E denotes work input (or output), W denotes mass flow, and h denotes enthalpy. The left-hand side of the equation is equal to the total energy crossing the control volume boundary as heat and work per unit time. The right-hand side is equal to the change in total energy of the control volume per unit of time (potential and kinetic energy is neglected). Neglecting any heat losses of the heat exchanger ($Q=0$), the left-hand side is equal to zero, thus,

$$W_g(h_{g,\mathrm{in}} - h_{g,\mathrm{out}}) = W_{\mathrm{H}_2}(h_{\mathrm{H}_2\mathrm{out}} - h_{\mathrm{H}_2\mathrm{in}}) \tag{2}$$

Knowing the mass flows, the inlet condition for the combustion gases (subscript g) as well as the inlet and the required outlet condition for the hydrogen for a certain operating point, the enthalpy of the gases at the outlet may be calculated and, thereby, also the temperature drop of the gas flow. The hydrogen inlet temperature is assumed to be 24 K. In accordance with the methodology for recuperators, a thermal effectiveness ε (based on the assumption of equal specific heats of the two streams) of the HE may be defined as the ratio of hydrogen temperature rise to the ideal value, the latter being the difference between the gas and hydrogen inlet air temperatures [7]

$$\epsilon = \frac{T_{\mathrm{H}_2\mathrm{out}} - T_{\mathrm{H}_2\mathrm{in}}}{T_{g,\mathrm{in}} - T_{\mathrm{H}_2\mathrm{in}}} \tag{3}$$

In the calculations, the effectiveness is computed for the design point using Eq. (3), and for off-design conditions, the temperature drop prior to the exhaust nozzle is calculated by assuming that the effectiveness is preserved. In practice, during part load conditions, the effectiveness increases as physical mass flow reduces while the volume or area remains fixed [7]. However, as the temperature drop in the exhaust has a small influence on the thrust output, the exclusion of increased effectiveness for decreased power settings will have a negligible effect on the off-design performance.

Knowing the required heat transfer Q to heat up the hydrogen as desired and the temperatures of the gas and hydrogen in and out of the heat exchanger, the area A, required to accomplish the heat exchange can be estimated. Note that this Q is different from the Q used in Eq. (1), which concerned heat fluxes across the boundaries of the control volume. The objective here is to estimate, approximately, the area requirement if using the struts to evaporate the hydrogen. From the definition of the overall heat transfer coefficient U, the heat transfer for a heat exchanger can be written as [8]

$$Q = UA\Delta T_m \tag{4}$$

where ΔT_m is the overall mean temperature difference for the heat exchanger. If effects of fouling are neglected and the overall surface efficiency is assumed equal to 1, the total thermal resistance to heat transfer between two fluids can be expressed as the sum of the thermal resistance for convection on the hot and cold sides, respectively, and the thermal resistance for conduction [8]

$$\frac{1}{UA} = \frac{1}{(\alpha A)_h} + \frac{b}{\lambda A_{\mathrm{HE}}} + \frac{1}{(\alpha A)_c} \tag{5}$$

where α is the convection heat transfer coefficient, b is the wall thickness (between the hot and cold fluids), λ is the thermal conductivity of the heat exchanger material, and subscripts h and c refer to the hot and cold sides of the heat exchanger (HE), respectively. The expression for ΔT_m is dependent on the directions of the flow streams within the heat exchanger, i.e., whether it is a parallel-flow or counterflow heat exchanger. In addition, it could be a mix of these types. As for the case discussed here to warm up the hydrogen, the heat exchanger appears to be more like a counterflow type. Following Incropera and DeWitt [8], the appropriate mean temperature difference for a heat exchanger is a log mean temperature difference. For a counterflow heat exchanger, the following expression for the log mean temperature difference is derived:

$$\Delta T_m = \Delta T_{lm} = \frac{(T_{h,\mathrm{out}} - T_{c,\mathrm{in}}) - (T_{h,\mathrm{in}} - T_{c,\mathrm{out}})}{ln\left[\dfrac{T_{h,\mathrm{out}} - T_{c,\mathrm{in}}}{T_{h,\mathrm{in}} - T_{c,\mathrm{out}}}\right]} \tag{6}$$

The heat transfer coefficient is estimated using the Nusselt number, defined as

$$\mathrm{Nu} \equiv \frac{\alpha L}{\lambda_f} \tag{7}$$

where L is a characteristic length of the heat transferring object and λ_f is the thermal conductivity of the fluid in question. Considering the option of employing the struts to evaporate the hydrogen, the heat transferring area is assumed to be a flat plate in parallel flow. According to Incropera and DeWitt [8], the Nusselt number can be shown to be a function only of a dimensionless length parameter, Reynolds number (Re) and Prandtl number (Pr). Assuming a flat plate in parallel flow and that the flow is turbulent, i.e., the Reynolds number of a strut is larger than the critical value of 5×10^5, an average Nusselt number can be determined by [8]

$$\overline{\mathrm{Nu}}_L = \mathrm{Pr}^{1/3}[0.037(\mathrm{Re}^{4/5} - 871)] \tag{8}$$

This equation is valid for $0.6 < \mathrm{Pr} < 60$ and $5 \times 10^5 < \mathrm{Re} < 10^8$.

In this context an approximate number on area requirement is sufficient. Therefore, a number of simplifying assumptions are made, in accordance with guidance from Sundén [9]. The thermal resistance for conduction (second term of the right-hand side of Eq. (5)) is expected to be small in comparison to the resistance terms for convection. Furthermore, the convection heat transfer coefficient on the cold side α_c, where the fluid is a mixture of liquid and gaseous hydrogen, is expected to be essentially higher than that on the hot side at which the fluid is a warm gas of high velocity. For these reasons, therefore, the thermal resistance is assumed only to be constituted of the thermal resistance for convection on the hot side. An area assessment based on these simplifications indicates a minimum area requirement, as the thermal resistances for conduction and convection on the cold side are neglected.

Table 1 Relevant heat exchanger conditions

Parameter	Value
Q (kW)	1403.5
$V_{g,\mathrm{out}}$ (m/s)	450.9
$T_{g,\mathrm{in}}$ (K)	789.7
$T_{g,\mathrm{out}}$ (K)	770.6
$T_{\mathrm{H_2in}}$ (K)	24
$T_{\mathrm{H_2out}}$ (K)	250

Using TURBOMATCH (Cranfield University's own software for gas turbine engine design/off-design performance calculations), the required heat transfer, the temperature of the gas flow at inlet and outlet, and the exhaust velocity are calculated for the design point for the V2527-A5 engine (as shown later, the operating point corresponds to the third column in Table 3). The version of TURBOMATCH employed here is changed to enable simulation of hydrogen-fueled engines [5]. Relevant heat exchanger conditions are listed in Table 1 and approximate dimensions of the exhaust struts are given in Table 2.

Assuming that the thermophysical properties of the exhaust gases are approximately equal to those of air at atmospheric pressure at the average gas temperature and taking the chord of the struts as the characteristic length, the flow turns out to be turbulent, i.e., the Reynolds number is larger than the critical value (5×10^5). Using Eqs. (8) and (7), the average convection heat transfer coefficient becomes 400.9 $\mathrm{W/m^2\,K}$. Equation (6) gives a log mean temperature difference, ΔT_{lm}, equal to 637.5 K. Combining Eqs. (4) and (5), a minimum heat-transferring area at the gas side equal to 5.5 $\mathrm{m^2}$ is obtained.

Considering the approximate strut dimensions, the total area on the gas side becomes ~ 0.82 $\mathrm{m^2}$, which is significantly less than the magnitude required to heat up the hydrogen from 24 K to 250 K. Thus, in order to employ the struts as the heat exchanger, an additional heat source, such as an external system, is needed. The simplified calculations outlined here suggest that the struts would be sufficient to heat up the hydrogen only by about 25–30 K. One option would, therefore, be to use an external heat source for the remaining temperature rise prior to the struts. An alternative is to place a heat exchanger in the exhaust, featuring a simple coil tube placed over the inside face of the jet pipe casing. Although this solution may incur additional pressure losses in the exhaust, it is, as discussed previously, believed that it would give a relatively aerodynamically clean jet pipe.

In all calculations presented in this paper, the energy necessary to raise the fuel temperature from the storage temperature is taken into account by decreasing the exhaust gas temperature accordingly. A detailed design of a heat exchanger is beyond the scope of this paper, and no reasonable figures on possible losses due to the

Table 2 Approximate dimensions of the V2500 exhaust struts [10]

Parameter	Value
Chord (mm)	173
Root radius, inlet (mm)	259
Root radius, outlet (mm)	230
Tip radius, inlet (mm)	461
Tip radius, outlet (mm)	422
Thickness (mm)	35
Lean (deg)	10
Number of struts	Unknown, ~12

Table 3 Performance comparison at takeoff conditions

	Kerosene	H_2 COT	Δ (%)	H_2 F_N	Δ (%)
W_1 (kg/s)	369.3	369.3	–	369.3	–
W_F (kg/s)	1.167	0.4369	−62.6	0.4136	−64.6
F_N (kN)	117.9	121.7	3.2	117.9	–
COT (K)	1555	1555	–	1522	−2.2
TET (K)	1471	1473	0.2	1443	−1.9
SFC (g/kN s)	9.8999	3.5909	−63.7	3.5092	−64.6
ESFC (J/N s)	426.9	430.9	0.9	421.1	−1.4
SPT (m/s)	319.2	329.5	3.2	319.2	–
A_{corenoz} (m^2)	0.3343	0.2862	−14.4	0.3228	−3.4
W_{outcore} (kg/s)	64.839	64.109	−1.1	64.086	−1.2
V_{outcore} (m/s)	387.8	450.9	16.3	391.9	1.1

inclusion of a heat exchanger are available. Therefore, no other effects of a heat exchanger than an exhaust gas temperature drop are taken into account in subsequent calculations. Investigating the design of a heat exchanger and its consequences on engine performance and weight is recommended as further work.

2.3 Engine Performance Estimations. The effects on the engine performance by changing to hydrogen fuel are estimated for the V2527-A5 engine using TURBOMATCH. Three engines with the same bypass ratio (BPR), overall pressure ratio (OPR), and inlet mass flow—one using kerosene and two using hydrogen—are simulated for the design point. When changing to hydrogen, either the combustor outlet temperature or the net thrust could be retained. Performance data for these two options together with the datum engine burning kerosene are shown in Table 3.

Because of the considerably higher heating value of hydrogen, the fuel flow to achieve the same combustor outlet temperature (COT) or net thrust is reduced by almost two-thirds. When the COT is preserved, the net thrust increases by 3.2% for this particular case, resulting in a corresponding increased specific thrust. The energy consumption, reflected here as energy specific fuel consumption (ESFC) (SFC times the lower heating value), increases slightly, but less than ~1%. Despite the fact that the mass flow through the core nozzle decreases, the velocity increases and a smaller nozzle area is required. The other option considered is to retain the net thrust. This is obtained by lowering the COT by 33 K. In terms of energy consumption, this is slightly beneficial, since ESFC decreases by 1.4%. The values of area, nozzle mass flow, and velocity of the bypass stream are not displayed, since they do not change when changing fuel.

Following Boggia and Jackson [4], the performance improvements could be explained by the two fundamental changes when using hydrogen: reduced mass flow and changed composition of the gases expanding through the turbine(s). Whereas the latter improves the performance, the former deteriorates the performance. Reduced mass flow through the turbine lowers the thrust output for two reasons. Decreasing the fuel flow implies that the exhaust mass flow decreases accordingly (see Table 3); hence, without any variation in gas composition, the thrust output decreases. For the computed cases, the mass flow through the core nozzle decreases by ~1%. In addition, a reduced mass flow through the turbine will result in a higher total temperature drop and, thereby, also a total pressure drop across the turbine in order to deliver the same amount of power to the compressor. As a consequence of the lower total temperature and pressure at the turbine exit, both the pressure thrust (thrust due to different pressure at engine inlet and exit) and the momentum thrust decrease. The latter is an effect of decreased core nozzle velocity.

However, the loss in thrust due to reduced mass flow is offset by the increased thrust owing to changed properties of the combustion products [4]. With the use of hydrogen, the combustion products contain no CO_2 and a larger portion of H_2O, which has a higher C_p value than CO_2. Having investigated a simple turbojet engine, Boggia and Jackson [4] concluded that the C_p value has increased by ~4% in the hot section of the engine when changing to hydrogen fuel use. Increased C_p value through the turbine will similarly, but in the opposite direction as reduced mass flow, affect the performance. For a fixed power output, it will cause smaller total temperature and pressure drops across the turbine. Provided that the core nozzle is not choked, a larger nozzle expansion ratio will result in a larger exhaust velocity, which, in turn, will increase the momentum thrust. In total, the positive effect of increased C_p value outweighs the effect of reduced mass flow and, hence, results in an increased net thrust when switching to hydrogen and retaining the COT.

Since ESFC does not decrease when changing to hydrogen and preserving the COT, slightly more fuel in terms of energy is required when burning hydrogen to attain a certain COT for the configuration investigated here. It should be pointed out that the energy consumption to attain a certain COT is highly dependent on the fuel-injection temperature and the location of the heat exchanger (HE) used to evaporate the liquid hydrogen. By heating the fuel more, it is possible to achieve performance benefits (see Sec. 2.4).

The effects on engine performance are quite small, but still there are some desired features that could be exploited. If the COT is kept the same, the turbine entry temperature (TET) is also about the same, thus requiring the same cooling technology. An increased specific thrust implies that the inlet mass flow, and thereby the physical size of the engine, may be reduced to achieve the same net thrust. The in-flight performance will, in turn, be beneficially affected through reduced engine drag and weight. However, to fully address the mission performance of a cryoplane compared with a conventional aircraft, the increased thrust requirement at cruise needs to be included (see [11]).

As for the option of lowering the COT to preserve the net thrust, there are a few interesting consequences worth mentioning. The decrease in TET of more than 30 K will require less advanced cooling technology as well as having a favorable effect on turbine blade life. The lowered ESFC will reduce the amount of energy that needs to be provided by the fuel for a certain thrust level. Moreover, designing for a lower maximum cycle temperature will help to suppress the NO_x emissions (see Sec. 3.1).

2.4 Unconventional Engine Cycles. As for the case of evaporating the hydrogen using a HE suited within the engine cycle, the cryogenic properties of hydrogen could be exploited to improve the performance through the usage of unconventional cycles. Some of these cycles are also proposed for kerosene-powered cycles; however, the main advantage when using hydrogen is the involvement of an additional fluid with favorable properties. Considering unconventional designs, there are basically four options, and various combinations of these that could be employed [2,4,5,12,13]:

- preheating the hydrogen fuel with exhaust gases
- cooling the compressor air with hydrogen fuel
- cooling turbine cooling air with hydrogen fuel
- hydrogen topping cycle

The design principle, objectives, advantages, and drawbacks of these concepts are outlined in Secs. 2.4.1–2.4.5.

2.4.1 Pre-heating the Hydrogen Fuel With Exhaust Gases. If the hydrogen was to be evaporated using a heat source within the engine cycle, a HE suited at the LPT exit is considered to be the safest solution. This configuration also would impose the minimum number of hardware changes. Using this configuration and heating the hydrogen to the minimum fuel temperature imposed by fuel system control needs was considered to be the "conventional" hydrogen engine configuration within the CRYOPLANE project [3]. However, by heating the fuel more, a larger amount of energy needs to be taken out of the exhaust flow, but on the other hand, less fuel would be required to achieve a certain COT, thereby providing performance benefits. In order to extract more energy from the exhaust flow, the HE needs to be designed for a higher thermal effectiveness, which implies that the physical size needs to be increased.

By repeating the calculations for the engine discussed in Sec. 2.3 (without re-optimizing the outer-fan pressure ratio) for a fuel temperature of 600 K instead of 250 K and preserving the COT, it was found that the ESFC decreased to 414.2 J/N s, the net thrust decreased to 120.7 kN, and the HE thermal effectiveness (Eq. (1)) needs to be increased from 0.295 to 0.755. Thus, at the expense of a significantly increased HE effectiveness and a slightly decreased net thrust, the energy consumption to attain the same COT may be decreased by 3.7% as compared to the kerosene-fueled engine. As hydrogen has a high C_p value, it is capable of capturing large amounts of energy, hence, reducing the ESFC [2]. The small thrust loss is explained by the increased temperature drop prior to the exhaust nozzle.

Although an effectiveness of 0.755 is not large compared to those of recuperators, which are designed for an effectiveness of about 0.8, it might not be reasonable to exclude pressure losses in the exhaust nozzle when evaluating the engine performance. If an external heat source is not employed, a heat exchanger suited at the LPT exit to increase the fuel temperature to 600 K might give rise to non-negligible pressure losses in the exhaust nozzle.

These results are in agreement with the results of Boggia et al. [2], who found that the SFC decreased by 3.9%, while the SPT decreased slightly, by 0.8%, when heating the fuel to 600 K instead of 250 K.

2.4.2 Cooling the Compressor Air With Hydrogen Fuel. As the compression work is roughly proportional to the temperature rise across the compressor (assuming constant C_p), cooling the compressor air offers performance advantages. If the compressor outlet temperature is fixed, the OPR could be increased, which mainly affects the SFC beneficially. Alternatively, the excessive work could be used to provide a higher fan pressure ratio, thus giving a higher fan thrust.

The compressor air could be cooled before the compressor (precooling), between the HP and LP compressor (intercooling), or continuously during the compression process. The latter is accomplished through the cooling of the stator and intermediate guide vanes along the compression process. This technique offers the highest performance gains, but technically, it is also the most complicated method to use [13].

Using a HE suited before the booster inlet of the V2527-A5 engine, Boggia et al. [2] found that the SFC decreased by 5.7% and the SPT increased by 6.7% compared to their reference cycle. To obtain these benefits, they did increase the OPR and BPR, and even when the increased weight (due to the heat exchanger) is included, the performance is improved because the thrust to weight ratio is increased by 1.2%.

There are, however, problems with designs featuring a HE at engine inlet. In addition to the increased engine weight and complexity owing to the inclusion of a HE, the main disadvantage of this concept is the risk of foreign object damage [3]. Another problem, which is particularly pronounced for precooled cycles, is the risk of icing on the air side of the HE. If the air is cooled below the equilibrium temperature at which the partial pressure of the water vapor is equal to its vapor pressure, water vapor will deposit out on the HE walls directly as an ice film [13]. Ice and frost formation initially tend to increase the heat transfer coefficient, but, eventually, clogging will occur and, consequently, the air flow will be reduced. Nevertheless, Payzer and Renninger [13] concluded that air-to-fuel HEs that fully justify the engine fuel flow without air side freezing can be accomplished.

2.4.3 Cooling Turbine Cooling Air With Hydrogen Fuel. Normally, the NGVs and the rotor of the HP turbine need to be cooled as the TET exceeds the maximum allowable metal temperature. This is accomplished by bleeding off a portion of the compressed air that is passed through cooling passages inside the blades. By cooling the bleed air using the hydrogen fuel in a hydrogen-to-air HE, the blade cooling is made more effective and, hence, the maximum allowable TET is increased. The energy output of the core is highly dependent on the level of TET. If it is increased, more energy needs to be provided through the fuel, but the thrust output increases significantly. By employing this concept (increasing TET by 140 K) and redesigning the cycle through increased BPR and number of LPT stages, Boggia et al. [2] attained an increased thrust to weight ratio of 8.4% and a decreased SFC of 2.1%.

One drawback associated with this technology is the complex design problems that have to be addressed when the routing of the turbine cooling air is redesigned. Another drawback is the severe thermal gradients in the blades due to the colder cooling air, which might cause low-cycle fatigue damage [12].

Although very advanced in terms of design features, it might in the very long run be possible to cool the relevant parts of the turbine directly using the hydrogen, without the usage of any HE. This would imply that there would be no decrease in cycle performance as a consequence of the normal cooling air extraction [13]. Moreover, the losses associated with the heat exchanging process are eliminated, and the introduction of any significant weight penalties is avoided.

2.4.4 Hydrogen Topping Cycle. In order to boost the net thrust output, the main engine could be mechanically coupled to a topping loop. The topping loop could, depending on how much increased engine complexity is accepted, be more or less complex. The configuration proposed by Boggia et al. [2] consists of a fuel pump, HE, compressor, combustor, and turbine. In their design, a fraction of the compressor air flow, which is fed to the topping cycle where it is precooled (in a hydrogen-to-air HE) and compressed, is bled off. Hydrogen is compressed, preheated (in the hydrogen-to-air HE), and burned very fuel rich in a combustor that is provided with the bled air. The combustion products, composed of a mixture of stoichiometric products and hydrogen, are then expanded in the topping turbine and fed to the main engine burner where the hydrogen burns completely with the main air stream. Using this configuration, it was found that the SPT increased by 11.1% and the SFC decreased by 1.1% compared to their reference cycle.

Looking at the practical feasibility of the additional components, Boggia et al. [2] found that the small dimensions of the radial turbomachinery to perform the topping cycle compression do not pose any design problems. Neither is the design of the combustor of the topping loop expected to cause any problems. However, they claim that the sealing of the topping cycle, where a fuel-rich, high-pressure hot mixture must be kept separate from the air, would be technically very difficult to achieve safely. Owing to this, the concept was rejected as impractical.

2.4.5 Conclusions on Unconventional Engine Design. According to Boggia et al. [2], who studied and compared the four different cycles described above, the precooled and high TET cycles appeared to be the most promising, offering a reduction in operational cost on the order of 3%. Furthermore, they point out that the positive aspect of these two cycles is that the innovations are technically feasible and do not involve any additional turbomachinery. The configuration with preheating of the fuel is the simplest solution and would require only minimum modifications to the base engine.

Payzer and Renninger [13] conclude that there is a definite thermodynamic benefit in switching from hydrocarbon fuel to hydrogen fuel because of the difference in fuel properties. However, having evaluated eight different unconventional cycles, both in terms of uninstalled engine performance and aircraft mission performance, they state that the additional complexity associated with these cycles does not appear to be justified.

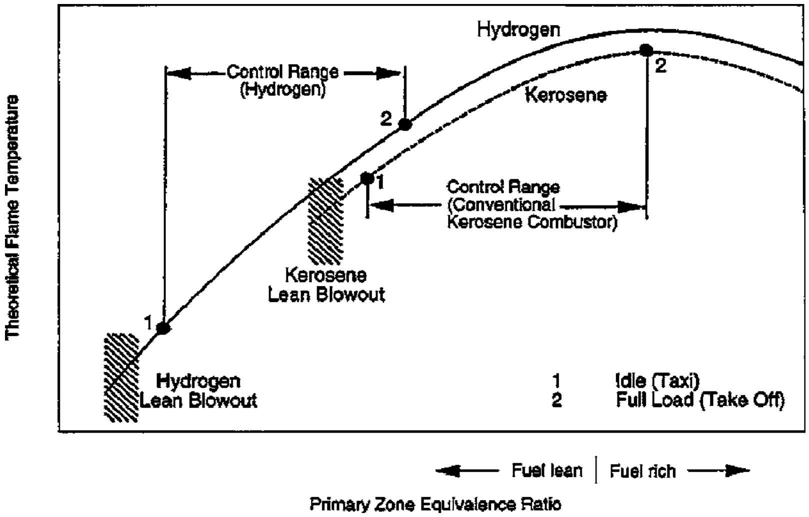


Fig. 2 Temperature characteristics of the combustion chamber primary zone [16]

3 Pollutant Emissions

When burning hydrogen, the emissions are reduced to water vapor (H_2O) and oxides of nitrogen (NO_x). In fact, the emission index of H_2O increases by a factor of ~2.6. All emissions containing carbon and sulphur are eliminated. Besides, for reasons discussed in this section, there is the potential for lowering the NO_x emissions as compared to conventional combustors using kerosene. Therefore, to achieve satisfactory emission characteristics, the challenge is to minimize the NO_x emissions while preserving the SFC. Performing research focused on reducing NO_x emissions of gas turbines is highly important, since the effects of these emissions on the atmosphere and on human health are considerable. At ground level, the presence of NO_x results in an increase in ozone concentration, which during prolonged exposure may cause respiratory illness, impaired vision, headaches, and allergies [14]. Moreover, NO_x emissions contribute to the greenhouse effect (via formation of ozone), cause damage to plant life, as well as add to the problem of acid rain. At high altitudes (above ~15 km) NO_x emissions cause ozone depletion and, hence, increased ground-level ultraviolet radiation, which might cause skin cancer and eye diseases [15].

This section is introduced by a brief discussion of the NO_x formation processes, followed by a section dealing with the consequences on the NO_x emissions of using hydrogen (Sec. 3.1). In Sec. 3.2, various possible hydrogen combustors are described. Finally, the potential of achieving low NO_x emissions is illustrated by conducting simulations for the V2527-A5 engine (Sec. 3.3).

3.1 NO_x Emissions. Most of the nitrogen oxides formed during combustion are in the form of NO; however, subsequently the NO oxidizes to NO_2. Usually these are lumped together, and the result is expressed in terms of oxides of nitrogen (NO_x). The formation processes of NO in combustion are complex and generally comprise different mechanisms: thermal NO, prompt NO, nitrous oxide (N_2O) mechanism and fuel NO [14]. Thermal NO formation is, for the hydrogen combustion, the most contributing mechanism to the total NO formation. Thermal NO is produced from the nitrogen and oxygen present in the air in the high-temperature regions of the flame and in the postflame gases. The process is endothermic and proceeds at a considerable rate, only at temperatures above around 1850 K [14]. Prompt NO is a designation for NO, which, under certain conditions, is formed very early in the flame region. Its initiationreaction includes the carbon atom. Prompt NO may play a significant role for lean systems burning kerosene, whereas it does not contribute to the NO formation in combustion systems using hydrogen (since there are no carbon atoms present in the fuel). In this context, considering high engine operating conditions at which most of the NO is formed, the nitrous mechanism is less important and is therefore omitted. Fuel NO is not relevant here, since neither kerosene nor hydrogen fuel contain any nitrogen.

In order to address properly the issue of NO_x production of hydrogen combustion compared to kerosene combustion, several different combustion characteristics of these fuels need to be considered. The engine load in a gas turbine is controlled by varying the TET, which is determined by the overall fuel-air ratio in the combustion chamber. When a conventional kerosene-fueled combustor is operating at full power, the primary zone operates at roughly stoichiometric fuel-air mixtures, where the flame temperature is highest and the chemical reactions are fastest [16]. At low power idle conditions, the fuel-air ratio is essentially leaner and the primary zone fuel-air ratio has to be maintained above the flame-out limit with a sufficient margin. Hence, in order to minimize the NO_x emissions, it is desirable to modify the fuel-air ratio in the combustor primary zone in a way that lean combustion is achieved at all engine load conditions without suffering a flameout [16]. In Fig. 2, the flame temperature of kerosene and hydrogen combustion versus primary zone equivalence ratio is shown.

The comparison illustrated in Fig. 2 applies only to a conventional kerosene-fueled combustor featuring a primary, intermediate, and dilution zone. If the kerosene-powered combustor would employ any unconventional technology, such as, for instance, a staged or an LPP (lean premixed prevaporized) combustor, the situation would change. Generally, the combustion zone of these combustors is operated at an equivalence ratio that is closer to the lean blowout limit than the conventional and, hence, it is not applicable to the situation sketched in the figure.

As illustrated in Fig. 2, the stoichiometric flame temperature of hydrogen/air flames is, in fact, higher than that of kerosene/air flames (~100 K higher), which would have a detrimental effect on the NO_x production. For a given equivalence ratio, the flame temperature is higher when burning hydrogen, which means that more NO_x would be generated. However, more importantly is that the hydrogen flame has a wider flammability range; particularly, the lean limit is much lower than that encountered for kerosene flames [16]. Therefore, the entire operating range may be shifted further into the lean region, with reduced NO_x emissions as a consequence.

Because of the rapid reaction kinetics of hydrogen/air flames, the burning velocity increases about eight times when changing to hydrogen [16]. The higher flame speed will result in a shorter combustor, hence, reduced cooling requirements. Owing to this, more air will be available for the combustion, hence, offering lower flame temperatures, which will reduce the NO_x emissions. Another effect of the increased burning velocity is that the actual dwell time in the hot flame zone decreases, which is beneficial for hindering NO_x formation.

The fact that hydrogen is a gaseous fuel while kerosene is liquid, is also an argument in favor of hydrogen. The reason is that for liquid fuels there is always the potential of near-stoichiometric combustion temperatures in local regions adjacent to the fuel drops, and, thereby, high NO_x formation, even if the average equivalence ratio throughout the combustion zone may be appreciably less than stoichiometric [14]. However, for increasing flame temperatures, the bulk flame temperature becomes closer to the stoichiometric value, and local conditions around the fuel drop have less influence on the overall combustion process and NO_x emissions. Hence, the difference in NO_x production decreases and, for equivalence ratios close to the stoichiometric value, the difference becomes negligibly small.

Moreover, the minimization of the NO_x emissions of hydrogen combustion is facilitated by the fact that no other secondary combustion products (e.g., CO, unburned hydrocarbons (UHC), soot) need to be regarded in the combustor design process.

From the preceding discussion, it may be concluded that, theoretically, there seems to be the potential to design a combustion system burning hydrogen that produces less NO_x emissions than a system burning kerosene. By calculation of the flame temperature at different relevant operating conditions, this statement is illustrated in Sec. 3.3. In addition, an experimental investigation that confirms the potential of achieving really low NO_x emissions when using hydrogen is outlined. However, prior to that, combustor configurations burning hydrogen are described.

3.2 Combustor Configurations Burning Hydrogen. In order to exploit the favorable characteristics of hydrogen combustion, the combustor needs to be redesigned, compared to the conventional designs using regular diffusive combustion. When using conventional combustors, with their limited number of fuel injection points, the mixing of fuel and hydrogen tends to be incomplete. Moreover, the large-scale hydrogen diffusion flames form layers of stoichiometric mixtures of high local temperature and viscosity in which the NO formation rate remains high and acts as barriers against the further progress of mixing [16]. Generally, when burning hydrogen the attempts to reduce the NO_x emissions are focused on lowering the flame temperature, eliminating hot spots from the reaction zone and reducing the duration and exposure in the flame region [16].

In 1992 during the third phase of the Euro-Québec Hydro-Hydrogen Pilot Project (EQHHPP), analytical modeling and experimental tests of low-NO_x combustors for aero engines were performed [16,17]. Various burner concepts, both employing the principle of premixing and nonpremixing, for lean hydrogen combustion were evaluated in a generic can-type combustor.

These studies and other preceding studies have shown that lean premixed combustion is undoubtedly superior to any combustion scheme without premixing in terms of temperature pattern uniformity and NO production. However, premixing implies the major drawback of premature burning and flashback danger, which may cause structural damage and compromise the operational reliability [17]. The risk of autoignition for premixed systems and the problems of large-scale hydrogen diffusion flames have led to the lean non-premixing concept of micromix combustion, which is based on miniaturized diffusive combustion. Employing this principle, imperfections in fuel/air mixing cannot be avoided; however, the local standard deviations from the nominal equivalence ratio are kept small [16].

The micromix combustor consists of a very large number (typically, more than 1000) diffusion flames uniformly distributed across the burner's main cross section, thereby minimizing the geometric size of the combustion zone [16,18,19]. Theoretically, the lowest NO_x production would be achieved by an infinite number of miniaturized flames. The minimum size is, however, restricted by the manufacturing cost and the combustion stability at engine idle conditions; the latter being deteriorated with reduced flame dimensions. In addition to minimizing the scale of the combustion zone, the micromix concept aims at optimally utilizing the available pressure loss (providing energy for the dissipative turbulent mixing process) in the combustion system to enhance the mixing process of hydrogen and air. Simultaneously minimizing the scale of combustion and maximizing the intensity of mixing will minimize the number and size of local stoichiometric flame regions, where the gas phase NO formation processes are most likely to occur. Because of the high flame speed and high reactivity of hydrogen compared to other fuels, the degree of miniaturization of hydrogen/air diffusion flames very well exceeds the possibilities of other fuels [16].

Table 4 Characteristics of kerosene and hydrogen

	MW (g/mol)	Flammability limits in air ($\%_{vol}$)	LHV (MJ/kg)	Stoichiometric FAR (-)
Kerosene	175	0.6–4.7	43.124	0.0683
Hydrogen	2.016	4.0–75.0	120.0	0.02916
Air	28.966	–	–	–

3.3 Illustration of the NO_x Reduction Potential. The theoretical advantages of achieving low NO_x emissions outlined above are here further illustrated by simulation of the V2527-A5 engine, considered to be provided with different combustion systems. The aim of this section is to quantify and compare the leanest possible attainable levels for the operation conditions takeoff and idle (assumed to be 7% power setting) at static sea level, and the flame temperatures associated with these, for kerosene and hydrogen combustion, respectively. Rather than attempting to design a combustion system, the qualitative differences in flame temperature associated with these different fuels are assessed. In reality, when designing a combustion system for an aero engine, the most challenging operating condition, such as relight at high altitude and low flight mach numbers, needs to be addressed [7].

The takeoff condition is modeled with the performance code TURBOMATCH. As the compressor maps provided in TURBOMATCH were not sufficient to model the idle condition, GASTURB [20] is employed to simulate this operating condition. The characteristics for kerosene and hydrogen used in the calculations are given in Table 4 [21,22]. The stoichiometric fuel-air ratios are calculated by setting up a chemical balance and assuming stoichiometry.

The flame temperatures are calculated using the NASA CEA code assuming chemical equilibrium [23,24]. As the dwell time in a gas turbine combustor is slightly too short for chemical equilibrium to be achieved, the flame temperatures computed this way tend to be slightly overestimated. Nevertheless, these figures qualitatively indicate differences in flame temperatures and, thereby, the NO_x levels that may be expected from the different combustion systems.

Under normal operation conditions, the maximum percentage of air entering the combustor that may be introduced in the primary zone is restricted by the flame-out limit of fuel/air mixture. Knowing the mass flows of fuel and air in the primary zone (PZ), the mass fraction of fuel Y_F can be calculated. As the flammability limits shown in Table 4 are given in volume fractions X_F, the calculated mass fraction needs to be transformed accordingly. This transformation can be accomplished using the molecular weights (MW) of the fuel and air with [25]

Table 5 Comparison of combustion characteristics

	Kerosene T/O	H_2, T/O	Kerosene Idle	H_2, Idle
F_N (kN)	117.9	117.9	8.3	8.3
COT (K)	1555	1522	812	807
W_3 (kg/s)	54.1	54.1	10.4	10.2
T_3 (K)	835.1	836.2	478.6	476.7
P_3 (kPa)	2882	2882	398	393
Maximum air to the primary zone during idle				
Air to PZ (%)	31	95	31	95
X_F ($\%_{\mathrm{vol}}$)	1.14	10.36	0.45	4.43
ϕ_{pz} (-)	1.02	0.28	0.40	0.111
$\mathrm{T}_{\mathrm{flame}}$ in PZ (K)	2613	1587	1438	825
Maximum air to the primary zone during T/O				
Air to PZ (%)	78	95	31	95
X_F ($\%_{\mathrm{vol}}$)	0.46	10.36	0.45	4.43
ϕ_{pz} (-)	0.40	0.28	0.40	0.111
T_{flame} in PZ (K)	1733	1587	1438	825

$$X_F = \frac{Y_F}{\mathrm{MW}_F\left(\frac{Y_F}{\mathrm{MW}_F} + \frac{Y_A}{\mathrm{MW}_A}\right)} \tag{9}$$

In order to avoid weak extinction at low power, Walsh and Fletcher [7] recommend that the equivalence ratio in the primary zone for a sea-level static maximum rating is 1.02 for a conventional kerosene-fueled combustor. According to the calculations performed here, this equivalence ratio is achieved by distributing ~31% of the air entering the combustor to the primary zone. An overview of the results and combustor inlet conditions, in terms of mass flow, temperature, and pressure, is presented in Table 5. At idle, the overall volume fraction of fuel would become less than the lean blowout limit of $0.6\%_{\mathrm{vol}}$. However, since there still would be local regions of fuel-air ratios sufficiently high to sustain a flame, it would be possible to run the combustor leaner than this limit without suffering a flameout. According to Walsh and Fletcher [7], the weak extinction limit is around an equivalence ratio of 0.25, which is well below the equivalence ratio of 0.40 derived here for the idle condition.

At takeoff condition, an equivalence ratio of 1.02 in the primary zone corresponds to a flame temperature of 2613 K, which is ~300 K higher than that given in Walsh and Fletcher [7]. Assuming chemical equilibrium, comparing the flame temperature derived here with the number given in Walsh and Fletcher [7], it is indicated that overestimated flame temperatures by ~10–15% might be expected for stoichiometric conditions burning kerosene. This deviation is, however, not expected to substantially change the conclusions, since the main purpose of this exercise is to compare the flame temperatures (and the expected $\mathrm{NO_x}$ emissions production) obtained when burning two different fuels, and the difference in flame temperature between the two fuels is not expected to be affected significantly by this deviation. However, as the rate of chemical kinetics and the combustor lengths of these fuels are different, it cannot be excluded that the difference in flame temperatures would change if the flow field and the chemical kinetics were taken into account. At takeoff condition with a prevailing flame temperature of 2613 K (or rather 2300 K), depending on the combustion system, it is likely that excessive amounts of thermal NO are formed.

In order to reduce the flame temperature at takeoff conditions, variable geometry combustors could be employed. Practically, this could be achieved by the use of variable-area swirlers to control the amount of air flowing into the combustion zone, variable air openings into the dilution zone, or a combination of these [14]. Using this technology, a large proportion of air is admitted in the primary zone at maximum power conditions to lower the flame temperature and provide adequate film-cooling air, and as the power is decreased, an increasing proportion of this air is diverted to the dilution zone.

Theoretically, the lowest flame temperature that may be achieved by employing variable geometry combustors at full power conditions for a kerosene-fueled combustion system is the temperature obtained by assuming a primary zone equivalence ratio equal to the weak extinction limit including a safety margin. This is the leanest possible fuel-air mixture at which the combustion zone may be operated using any novel combustion concept. In order to achieve this condition in the combustion zone, there might be other novel combustor concepts more appropriate than a variable geometry combustor. According to the computations conducted for idle condition when the equivalence ratio at takeoff is equal to 1.02, this corresponds to an equivalence ratio of 0.40. An equivalence ratio of 0.40 is achieved at takeoff by diverting up to 78% of the incoming air to the primary zone, giving a substantially reduced flame temperature of 1733 K (see Table 5).

All forms of variable geometry systems are associated with drawbacks, like complex control and feedback mechanisms, which tend to increase cost and weight and reduce reliability [14]. Furthermore, Lefebvre [14] states that problems of achieving the desired temperature pattern in the combustor efflux gases could be encountered, especially if the liner pressure drop is allowed to vary too much.

As for hydrogen combustion when employing the concept of micromix combustion, there is no splitting of the combustor air into different zones. All available air, which is the total compressor exit mass flow minus bleed air and cooling air, is used for combustion. Considering the V2500 engine, the air flow at compressor exit is divided according to the following proportions [26]: 71.9% combustion air, 13.4% liner cooling air, and 14.7% bleed air. As a consequence of the high flame speed of hydrogen flames, it has been shown experimentally that the micromix combustor needs roughly one-third of the conventional combustor's liner length [18,26]. Accordingly, the liner cooling air may be reduced by a factor of about one-third, as an order of magnitude. Comparing to the numbers stated for the conventional combustor of the V2500 engine, ~5% of the air entering the combustor (having subtracted the bleed air) is needed for liner cooling of the micromix combustor.

Because of the very wide flammability limits of hydrogen/air flames, it would be possible to devote all air entering the combustor for combustion. However, owing to the cooling requirement, maximally 95% could be used for combustion. Allowing 95% of the air to take part in the combustion, a very lean fuel-air ratio is achieved (equivalence ratio equal to 0.28) in the primary zone for takeoff condition. This equivalence ratio is yielded both when maximizing the amount of air to the primary zone at takeoff and at idle. Despite the fact that the stoichiometric flame temperature is $\sim$100 K higher for hydrogen/air flames, the flame temperatures corresponding to the same operating condition are essentially lower than those yielded for the kerosene case.

Looking at the takeoff condition where most of the NO_x is formed, the flame temperature is 1587 K, which is $\sim$1000 K lower than the corresponding figure for kerosene. At idle, the relation is in the same order of magnitude. If the kerosene combustor would be provided with a variable geometry combustor, theoretically the difference might be reduced to $\sim$150 K. As the hydrogen-fueled combustor would be operated with a remarkably larger margin to the weak extinction limit than would the kerosene-fueled combustor provided with a variable geometry combustor, the hydrogen-fueled combustor is likely to be the more viable of these two (see Sec. 4). The differences in flame temperatures will have a substantial effect on the amount of thermal NO produced by the combustion system. In addition, since the lean kerosene-powered combustor would probably also generate prompt NO, whereas the combustor burning hydrogen would not, the difference in NO_x formation between these two fuels is likely to be even larger than assessed when looking at the primary zone flame temperatures. Even if the flame temperatures of the hydrogen combustor are low for both takeoff and idle, they are higher than the lower limit, which is the temperature required by the cycle needs to give the desired performance, namely, the combustor outlet temperature (COT).

In spite of the simplicity in the approach of the study, the results of the calculations clearly suggest that there is the potential to design a combustion system using hydrogen that produces less NO_x emissions than any system burning kerosene. The NO_x reduction potential when burning hydrogen has also been shown experimentally by providing the Honeywell APU GTCP36-300 engine with various combustion systems, including the present configuration of the micromix combustor, and measuring the NO_x mole fractions [18]. The authors state that significantly lower levels of NO_x may be achieved with the micromix combustors compared to conventional combustors using kerosene. For main engine start conditions, the mole fractions of NO_x emissions are reduced by $\sim$88% compared to the minimum change configuration (only hydrogen injection nozzles), and by roughly 86% when compared to the emissions of the kerosene-fueled engine. As the output power is reduced, the difference increases. There is, however, no use in simply changing the injection nozzle of the original combustor to hydrogen injection nozzles; this would even increase the NO_x emissions.

4 Design and Handling Issues

In order to adapt the combustion chamber for hydrogen, some changes are necessary and some others are desired in order to utilise the favorable properties of hydrogen in an optimal fashion. The minimum change that has to be done to adapt a conventional combustion chamber for hydrogen, is to replace the injection system because, when the hydrogen is injected into the combustor, it is in the gaseous state, while the kerosene is liquid. In addition, there are several other changes that need to be considered to optimally utilize the changed conditions owing to burning hydrogen. The higher flame speed will result in a shorter combustor and, hence, reduced engine weight and combustor liner cooling requirements.

When burning hydrogen, the thermal energy radiated to the surroundings is lower than that of kerosene, thereby beneficially affecting the liner durability and liner cooling requirements. The reason for the reduced radiation could be explained as follows. In most gas turbine combustors, a sizeable proportion of the heat from the hot gases contained within the combustor to the liner wall is by radiation. Generally for combustion gases generated by combustion of kerosene, the total emitted radiation comprises two components [27]: “nonluminous” radiation, which emanates from certain heteropolar gases (notably, CO_2 and H_2O), and “luminous” radiation, which depends on the number and size of the solid particles (mainly soot) in the flame. At high levels of pressure encountered in modern gas turbine combustors, the flame is characterized by a predominance of luminous radiation [27]. Since no particles at all are present in the combustion gases when burning hydrogen, there is no luminous radiation, thus essentially lowering the total heat radiation as compared to combustion gases of kerosene combustion.

As described above, hydrogen/air flames have an essentially wider flammability range, which will allow operation of the combustion zone at an equivalence ratio that has a large margin to the lean blowout limit. This will facilitate the handling of the engine as well as reduce the creation of white noise (noise with lots of different frequencies) that might give rise to pulsations and vibrations in the engine, which, in turn, via resonance can have a detrimental effect on engine components.

5 Conclusions

From a technical point of view, it seems to be feasible to use hydrogen for aero gas turbines. The main changes comprise redesign of the combustion chamber and fuel control system, as well as the implementation of facilities to evaporate the hydrogen prior to its entry into the combustion chamber. The fuel heating can be accomplished either by an external heat source or a HE located at a suitable engine location. In practice, the heat exchanging might be accomplished by employing the struts, which are the mechanical structures in the exhaust which hold the rotors in place and are connecting the outer structure of the bearings. Alternatively, the HE could feature a simple coil tube placed over the inside face of the jet pipe casing, avoiding major engine changes and giving a relatively aerodynamically clean jet pipe. Simplified calculations suggest that the heat transferring area available from the exhaust struts is sufficient to accomplish only a minor portion ($\sim$10%) of the desired temperature rise of the hydrogen fuel.

Small engine performance gains, which depend on the fuel temperature and cycle configuration, in the order of a few percent may be obtained by changing to hydrogen fuel. In order to obtain a certain performance, a lower TET could be used, thus requiring less advanced cooling technology as well as having a favourable effect on turbine blade life. By employing unconventional cycles, it would be possible to increase these performance gains. However, it appears to be questionable if these benefits justify the increased complexity imposed by unconventional cycles.

In terms of pollutant emissions, hydrogen use offers the possibility of a significantly reduced number of emission species, resulting in only H_2O and NO_x emissions. In addition, the results of the calculations suggest that there is the potential to design a combustion system using hydrogen that produces less NO_x emissions than any system burning kerosene. With respect to combustor configurations burning hydrogen, the lean non-premixing concept of micromix combustion, which is based on miniaturized diffusive combustion, is suggested as a promising configuration.

Because of the wider flammability range of the hydrogen/air flames (compared to kerosene/air flames), the combustion zone can be operated at an equivalence ratio that has a larger margin to the lean blowout limit. As a consequence, the engine handling is facilitated and the creation of white noise is reduced. Furthermore, when burning hydrogen the thermal energy radiated to the surroundings is lower than that of kerosene, thereby beneficially affecting the liner durability and liner cooling requirements.

Acknowledgment

The authors would like to express their gratitude to Professor Friedemann Suttrop at Fachhochschule Aachen, Stefano Boggia at MTU Aero Engines, Anthony Jackson at Cranfield University, and Professor Bengt Sundén at Lund Institute of Technology for their advice and comments. Furthermore, the Swedish Energy Agency and the Swedish Defence Research Agency that funded this work are acknowledged. The calculations presented in the paper are performed using TURBOMATCH (Cranfield University's own software), GASTURB and the NASA CEA code.

Nomenclature

Abbreviations and Chemical Formulas

APU = auxiliary power unit
CEA = chemical equilibrium and applications
CO = carbon monoxide
CO_2 = carbon dioxide
EC = European Commission
EQHHPP = Euro-Québec Hydro-Hydrogen Pilot Project
H_2 = gaseous hydrogen
HE = heat exchanger
H_2O = water vapor
HP = high pressure
HPC = high-pressure compressor
ISA = international standard atmosphere
LP = low pressure
LPP = lean premixed prevaporized
LPT = low-pressure turbine
NASA = National Aeronautics and Space Administration
NGV = nozzle guide vane
N_2O = nitrous oxide
NO = nitric oxide
NO_2 = nitrogen dioxide
NO_x = nitrogen oxides ($NO+NO_2$)
PZ = primary zone
T/O = takeoff
UHC = unburned hydrocarbons

Notations

A = area (m^2)
b = wall thickness (m)
BPR = bypass ratio
COT = combustor outlet temperature (K)
C_p = specific heat at constant pressure (J/kg K)
E = work output (or input) (W)
ESFC = energy specific fuel consumption (J/N s)
F = thrust (N)
FAR = fuel-air ratio
h = enthalpy (kJ/kg)
L = characteristic length (m)
LHV = lower heating value (MJ/kg)
MW = molecular weight (g/mol)
Nu = Nusselt number
OPR = overall pressure ratio
P = pressure (kPa)
Pr = Prandtl number
Q = heat transfer (W)
SFC = specific fuel consumption (g/kN s)
SPT = specific thrust (m/s)
T = temperature (K)
TET = turbine entry temperature (K)
U = overall heat transfer coefficient (W/m^2 K)
V = velocity (m/s)
W = mass flow (kg/s)
X = volume fraction
Y = mass fraction
α = convection heat transfer coefficient (W/m^2 K)
Δ = change (any unit)
ε = thermal effectiveness of a HE
λ = thermal conductivity (W/m K)
ϕ = equivalence ratio

Subscripts

1 = engine intake front flange
3 = combustor inlet
A = air
c = cold
corenoz = nozzle of the core
F = fuel
f = fluid
g = combustion gases
h = hot
H_2 = hydrogen
HE = heat exchanger
L = characteristic length
lm = log mean
m = mean
N = net
outcore = exhaust of the core
pz = primary zone
vol = volume

References

[1] Rogers, H. L., Lee, D. S., Raper, D. W., Foster, P. M. de F., Wilson, C. W., and Newton, P. J., 2002, "The Impacts of Aviation on the Atmosphere," Aeronaut. J., **106**(1064), pp. 521–546.

[2] Boggia, S., Jackson, A., and Singh, R., 2001, "Unconventional Cycles for Aero Gas Turbine Engines Burning Hydrogen," *15th Symposium on Air Breathing Engines*, ISABE, Bangalore, India, September 3–7.

[3] Corchero, G., and Montañes, J. L., 2003, "An Approach to the Use of Hydrogen in Actual Commercial Aircraft Engines," *16th Symposium on Air Breathing Engines*, ISABE, Cleveland, OH, August 31–September 5.

[4] Boggia, S., and Jackson, A., 2002, "Some Unconventional Aero Gas Turbines Using Hydrogen Fuel," *Proceedings of ASME Turbo Expo 2002*, Amsterdam, ASME, New York, Vol. 2B, pp. 683–690.

[5] Boggia, S., 2001, "Four Unconventional Aero Gas Turbine Engines Burning Hydrogen - Cryoplane Project," M.Sc. thesis, Cranfield University, UK.

[6] Cengel, Y. A., and Boles, A., 1996, *Thermodynamics, An Engineering Approach*, McGraw-Hill, New York.

[7] Walsh, P. P., and Fletcher, P., 1998, *Gas Turbine Performance*, Blackwell Science, Malden.

[8] Incropera, F., and DeWitt, D., 1996, *Fundamentals of Heat and Mass Transfer*, Fourth ed., Wiley, New York.

[9] Sundén, B., 2004, Professor at Department of Heat and Power Engineering, Lund Institute of Technology, Lund University, Sweden, private communication.

[10] Jackson, A., 2004, Consultant at Cranfield University, UK, private communication.

[11] Svensson, F., Hasselrot, A., and Moldanova, J., 2004, "Reduced Environmental Impact by Lowered Cruise Altitude for Liquid Hydrogen-Fuelled Aircraft," Aerosp. Sci. Technol., **8**(4), pp. 307–320.

[12] Baerst, C. F., and Riple, J. C., 1979, "Preliminary Studies of a Turbofan Engine and Fuel System for Use With Liquid Hydrogen," Hydrogen in Air Transportation, International DGLR/DFVLR Symposium, September 11–14, Germany, sequence 19.

[13] Payzer, R. J., and Renninger, S. W., 1979, "Hydrogen Fueled High Bypass Turbofans in Subsonic Aircraft," Hydrogen in Air Transportation, International DGLR/DFVLR Symposium, September 11–14, Germany.

[14] Lefebvre, A. H., 1998, *Gas Turbine Combustion*, Second ed., Edwards Brothers, Ann Arbor.

[15] Singh, R., 2001, "An Overview: Gas Turbine Generated Pollutants and the Emerging Technology Solutions," lecture notes of course in gas turbine performance, June 11–15, Cranfield University, UK.

[16] Ziemann, J., Mayr, A., Anagnostou, A., Suttrop, F., Lowe, M., Bagheri, S. A., and Nitsche, Th., 1998, "Potential Use of Hydrogen in Air Propulsion," EQHHPP, Phase III.0-3, Final Report, submitted to the European Union.

[17] Ziemann, J., Shum, F., Moore, M., Kluyskens, D., Thomaier, D., Zarzalis, N., and Eberius, H., 1998, "Low-NO_x Combustors for Hydrogen Fueled Aero Engine," Int. J. Hydrogen Energy, **23**(4), pp. 281–288.

[18] Dahl, G., and Suttrop, F., 2001, "Combustion Chamber and Emissions, The Micromix Hydrogen Combustor Technology," Task Technical Report No. 4.4-5A, CRYOPLANE Project.

[19] Dahl, G., and Suttrop, F., 1998, "Engine Control and Low-NO_x Combustion for Hydrogen Fuelled Aircraft Gas Turbines," Int. J. Hydrogen Energy, **23**(8), pp. 695–704.

[20] Kurzke, J., 1998, *User's Manual - GasTurb 8.0 for Windows, A Program to Calculate Design and Off-Design Performance of Gas Turbines*, www.gasturb.de

[21] Goodger, E. M., 1993, *Hydrocarbon Fuel Chemistry, With Particular Reference To Aviation*, Landfall Press, Norwich, England.

[22] Wester, L., 1996, *Tabeller och diagram för energitekniska beräkningar*, Västerås, Sweden.

[23] McBride, B. J., and Gordon, S., 1996, "Computer Program for Calculation of Complex Chemical Equilibrium Compositions and Applications - II. Users Manual and Program Description," NASA Ref. Publ. No. 1311, from http://www.grc.nasa.gov/WWW/CEAWeb/

[24] Gordon, S., and McBride, B. J., 1994, "Computer Program for Calculation of Complex Chemical Equilibrium Compositions and Applications - I. Analysis," NASA Ref. Publ. No. 1311, from http://www.grc.nasa.gov/WWW/CEAWeb/

[25] Turns, R. S., 1996, *An Introduction to Combustion: Concepts and Applications*, McGraw-Hill, New York.

[26] Suttrop, F., 2003, Professor at Fachhochschule Aachen, Germany, private communication.

[27] Singh, R., 2002, "Combustor Cooling and Metal Temperatures," lecture notes of course in gas turbine combustion, June 24–28, Cranfield University, UK.

K. Kitagawa
Research Center for Advanced Energy Conversion,
Nagoya University,
Furo-Cho, Chikusa-ku,
Nagoya 464-8603, Japan
e-mail: kuni@apchem.nagoya-u.ac.jp

S. Itoh
Department of Applied Chemistry,
Graduate School of Engineering,
Nagoya University,
Furo-cho, Chikusa-ky,
Nagoya 464-8603, Japan

N. Arai
Research Center for Advanced Energy Conversion,
Nagoya University,
Furo-Cho, Chikusa-ku,
Nagoya 464-8603, Japan

Ashwani K. Gupta[1]
Department of Mechanical Engineering,
University of Maryland,
College Park, MD 20742
e-mail: akgupta@eng.umd.edu

Profiling of Redox Atmosphere in Flames by Chemical Seeding/Planar Laser-Induced Fluorescence (CS/PLIF)

Knowledge on the local value of reducing and oxidizing (redox) atmospheres in flames is among the most important issues to be desired by combustion engineers. In this study, the spatial distribution of a redox atmosphere in flames has been measured experimentally by the chemical seeding/laser-induced fluorescence (CS/LIF) technique. A solution of iron was sprayed into a premixed propane-air flame supported on a slot burner. The LIF intensity of FeO band was compared to that of a Fe line to estimate the experimentally determined degree of atomization in the reaction $FeO \rightarrow Fe + O$. *The flame temperature profile was determined as a rotational temperature and was obtained by comparing the LIF (laser-induced fluorescence) intensities of OH rotational lines. The degree of atomization was theoretically calculated on the basis that simple thermal dissociation takes place in the reaction. The redox atmosphere, or a redox index, is defined as the ratio of the experimentally determined to theoretically calculated degrees of atomization. Two-dimensional distributions or profiles of the excitation temperature, experimentally determined degree of atomization, and redox index have been measured using a charge coupled device (CCD) camera fitted with an optical bandpass filter and the associated signal processing using a computer. This method has been successfully applied to quantitatively illustrate the local atmosphere and profile of the redox atmosphere in flames.* [DOI: 10.1115/1.2179078]

Introduction

A variety of physical properties have to be measured in flames using key diagnostics in combustion systems. Temperature and flow velocity are among the important properties. However, these properties provide no information on the fate of the local chemical environment in the flame. It often becomes an important issue to determine whether the flame has a reducing or oxidizing (*redox*) environment in the material processing, such as for the processing of steel, other metals, and ceramics. The material surface is greatly affected by the combustion atmosphere. Furthermore, the atmosphere in flames is not uniform and often contains spatial distributions in practical flames. It is very difficult, if not impossible, to illustrate the local environment by theoretical modeling since a large number of chemical species provide the reducing and/or oxidizing role in complex systems.

In our previous study [1], the *redox index* was introduced and experimentally determined by a *chemical seeding* technique. In this technique, a chromium solution was continuously sprayed into a premixed acetylene-air flat flame. The flame temperature was experimentally determined as the excitation temperature by comparing spontaneous emission intensities of two chromium triplets. Chromium exists mostly as atoms or molecules of monoxide in acetylene flames [2]. The degree of atomization from chromium monoxide to atomic chromium was determined by comparing the spectral intensities of spontaneous emission from chromium monoxide and chromium atoms and by using the excitation temperature. The degree of atomization could be changed, depending on the temperature and reducing and oxidizing (*redox*), atmosphere in the flame. In order to extract the net effect by the latter, or to remove the temperature effect, the experimentally determined degree of atomization was normalized by the theoretically calculated value and was based on the simple thermal dissociation of chromium monoxide. The redox index thus obtained gave a measure of the oxidative, neutral, and reductive field in flames when the measured parameter in the combustion zone was smaller, equal, and greater than unity, respectively.

However, this method, based on chemical seeding/spontaneous atomic-molecular emission spectrometry (*CS/AMS*), suffers from the following disadvantages. Since the electronic excitation of atomic and molecular species of chromium requires high temperatures, the application of this technique is limited to flames at high temperatures, above 2000 K. In addition, the laborious and erroneous Abel inversion is indispensable to obtain the spatial distribution from the measured profile of spontaneous emission intensity. Moreover, the chromium seeding may cause toxic emission to the environment and, as such, is definitely improper for applications to practical combustion system, such as industrial furnaces and power plants. Iron is clearly more suitable as a seed element because of its negligible toxicity in various chemical forms. In iron/steel industry furnaces, it even exists without seeding the flow to the furnace. However, the drawback of using iron as the chemical seed is that the thermal excitation of iron atoms is more difficult than that of chromium atoms.

In order to overcome the problems described above, a new technique has been developed here that combines the *Fe-chemical seeding* and *planar laser-induced fluorescence (Fe-CS/PLIF)*. The technique has been successfully applied to quantitatively illustrate the spatial distribution of the redox atmosphere in a premixed propane-air flame at a low temperature. The flame temperature profiles have been measured with a two-color PLIF system.

---

[1] To whom correspondence should be addressed.

Contributed by the Combustion Committee of the International Gas Turbine Institute of ASME for publication in the Journal of Engineering for Gas Turbines and Power. Manuscript received November 22, 2002; final manuscript received September 15, 2005. Review conducted by S. Gollahalli.

Table 1 Spectral lines and partition functions

Wavelength (nm)	Electronic level energy (cm^{-1})	Ref.
FeI 293.6903[a]	0.000000–34039.513	7
FeI 368.7456[b]	6928.266–34039.513	7
FeI 368.7456[b]	7376.760–34039.513	7
FeO (1,0) 579.4[a]	0.0000–17292	2
FeO (0,0) 605.2[b]	0.0000–17292	2
OH $R_2(5)281.7336$[a]	0.0000–32682	3
OH $Q_1(10)284.4128$[a]	0.0000–32682	3

[a] Excitation line

[b] Main fluorescence line

$Z_{e,\mathrm{O}}(T)=(7.6413+7.4904)(T/10^3)-(2.0133\times10^{-1})(T/10^3)^2+(2.6166\times10^{-2})(T/10^3)^3-(1.2265\times10^{-3})(T/10^3)^4$

$Z_{e,\mathrm{Fe}}(T)=(1.0658+7.3013)(T/10^3)-2.2102(T/10^3)^2+(4.5301\times10^{-1})(T/10^3)^3-(4.0732\times10^{-2})(T/10^3)^4+(1.5017\times10^{-3})\times(T/10^3)^5$ Ref. 4

$Z_t=\{(2\pi mk_BT)^{1/2}/h\}^3 Z_v=1/1-\exp(-h\nu/k_BT)\exp(-h\nu/2k_BT) Z_r=8\pi^2 I_M k_BT/\sigma h^2$ Ref. 6

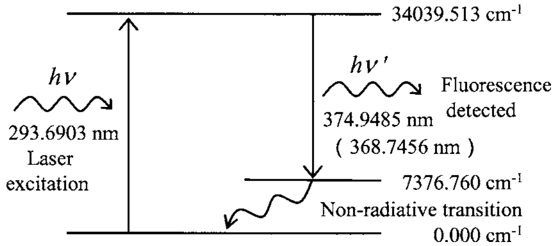


Fig. 1 Three-level model of atomic iron LIF

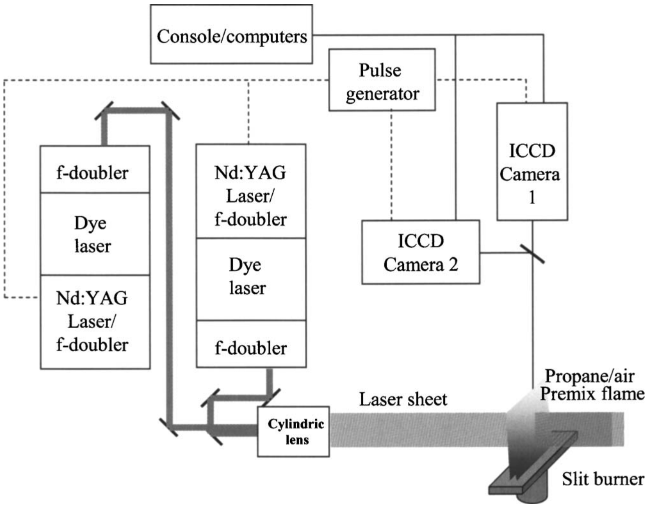


Fig. 2 Block diagram of experimental system

Theoretical

The spectral intensity of the LIF signal, S_f detected by a pixel of an intensified charge coupled device (ICCD) camera is given by the following simplified equation [3]:

$$S_f = KI_sB_LNg_L\exp(-E_L/kT)/QZ(T) \qquad (1)$$

where K is the instrument factor that involves the dependence of the pixel sensitivity on the wavelength; I_s the spectral intensity of the light source, or the laser; B_L, the Einstein's transition probability for absorption; N_T, the number density of species of interest; g_L, the statistical weight for the lower energy level to be excited by the laser; E_L, the energy of the lower level; k, the Boltzmann constant; T, the excitation temperature; Q, the quenching factor; and $Z(T)$, the partition function of the energy system of interest.

In order to simplify the analysis we make assumptions of identical excitation temperatures in the systems of iron atoms and FeO molecules. We also assume that the ratio of the quenching factors for these systems is independent of the temperature. Taking a ratio of the atomic fluorescence intensity R of Fe to the molecular one of FeO, we obtain the following equation:

$$\begin{aligned} R &\equiv S_{\mathrm{Fe}}/S_{\mathrm{FeO}} \\ &= C_R\left\{\frac{\beta_{\exp}g_{\mathrm{Fe}}B_{\mathrm{Cr}}Z_{e,\mathrm{FeO}}(T)}{(1-\beta_{\exp})g_{\mathrm{FeO}}B_{\mathrm{FeO}}Z_{e,\mathrm{Fe}}(T)}\right\}\exp\left\{\frac{(E_{\mathrm{FeO}}-E_{\mathrm{Fe}})}{kT}\right\} \end{aligned} \qquad (2)$$

Table 2 Instrumental specifications and conditions

Flame	Flat flame of premixed propane-air
Slot burner	Hitachi 207 AAS with a slit of 1.3 mm in width and 100 mm in length
Flow rate of C_3H_8	0.33, 0.40, and 0.51 L min^{-1}
Flow rate of air	10.1, 10.0, and 9.9 L min^{-1}
Equivalence ratio	0.80, 1.00, and 1.29
Chemical seeding	$Fe(NO_3)_3$ solution of 8.2×10^{-3} mol/L
Spray uptake rate	2.0 mL min^{-1}
Spray efficiency	ca. 1.5%
Two-color PLIF system	Tokyo Instruments Co., Combustion Analyzer
YAG laser	Continuum Powerlite 8000, 10 ns and 10 pps
Dye laser	Continuum ND6000, Rhodamine 590 and 610, 0.8–1.0 mJ/pulse
ICCD camera	Oriel InstaSpecV, 200 time integration for 20 c
Optical bandpass filter	302–400, 600–800, and 305–400 nm for Fe, FeO, and OH fluorescence measurements, respectively
Frequency doubler	Continuum UVT Generation
Monochromator	Acton Research Co., Specta-5

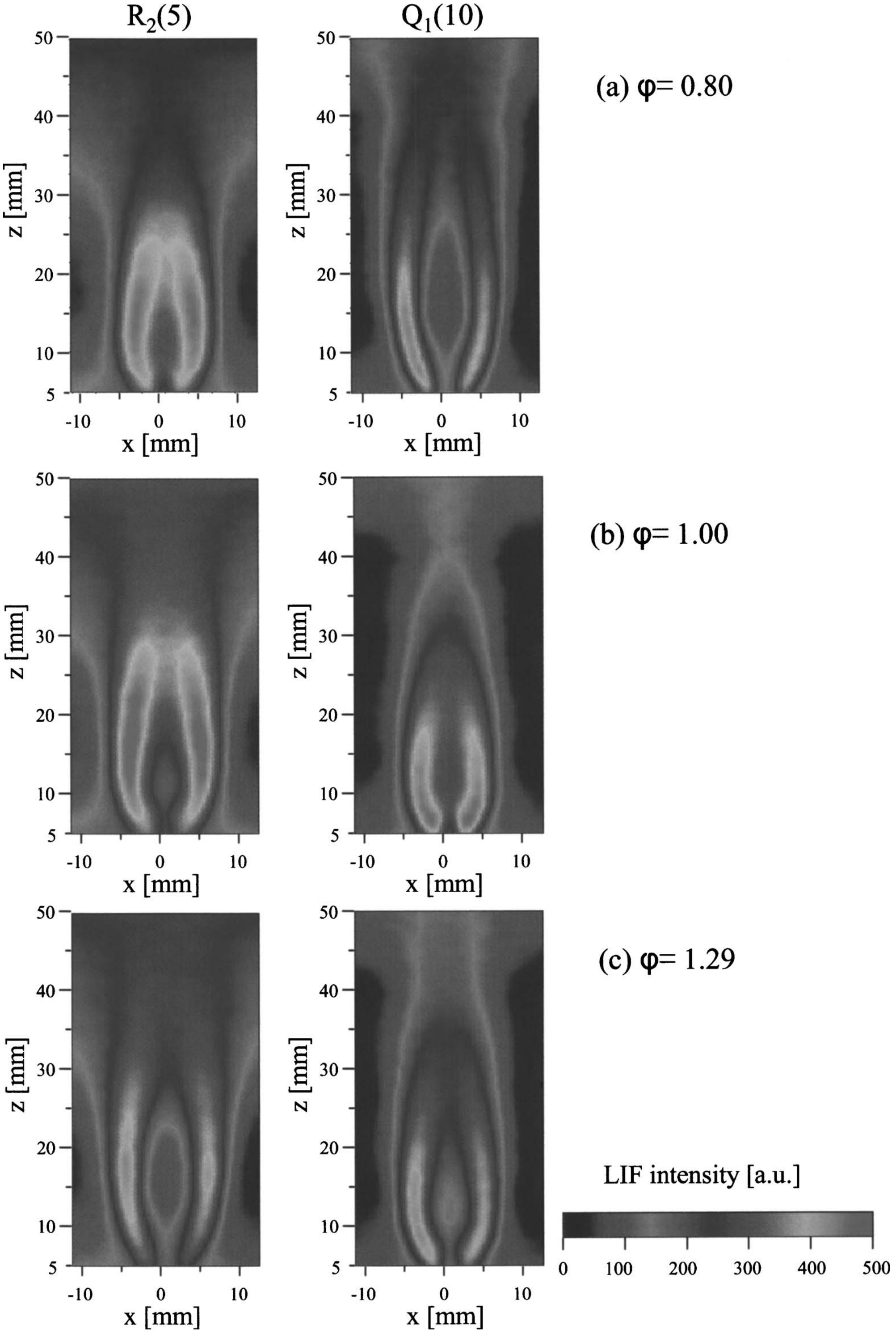


Fig. 3 Profile of LIF intensity of OH rotational line

where C_R is the constant involving the sensitivity ratio, depending on the wavelengths of the Fe line and the FeO band and the quenching factors; β_{exp}, the degree of atomization for $\mathrm{FeO} \rightarrow \mathrm{Fe} + \mathrm{O}$; Z_e, the electronic partition function; and the subscripts Fe and FeO stand for the species of interest. The levels of Fe and FeO to be excited by the laser are the ground states; the exponential term in Eq. (2) can be replaced by unity. The atomic partition of Fe has been given in Ref. [4] and that of FeO was derived from the definition, or the state sum, using the first two excitation levels given in Ref. [5]. After the excitation temperature has been determined by two-color PLIF of the OH rotational band, the partition functions can be estimated using the excitation temperature. All the other parameters are constant. Thus, the degree of atomization can be experimentally determined by the LIF intensity ratio of atomic and molecular species of iron.

On the other hand, the degree of atomization can be theoretically calculated as follows from the equilibrium constant, $K(T)$ on the basis of simple thermal dissociation [6].

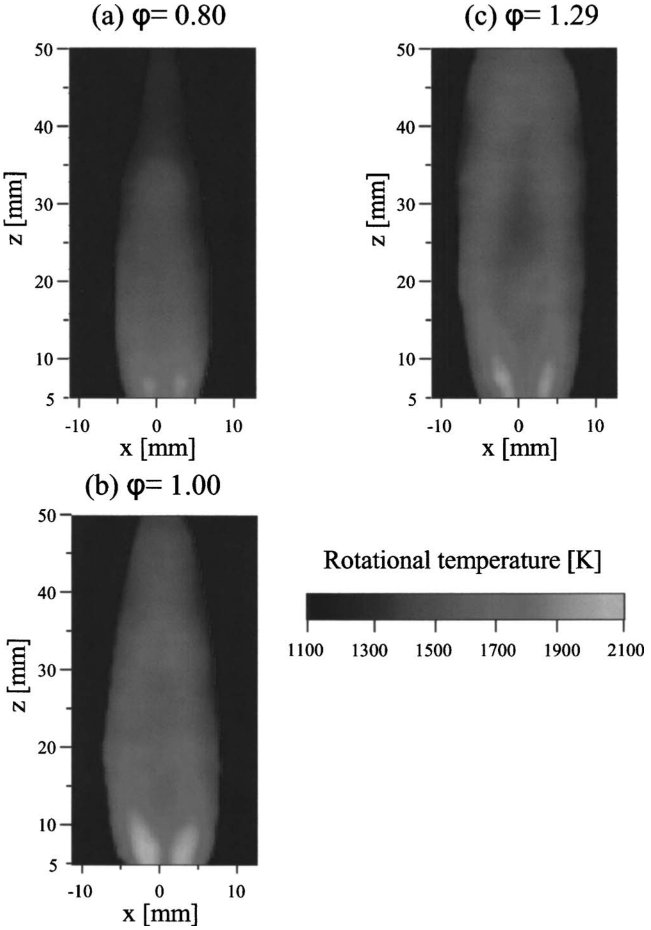


Fig. 4 Profile of rotational temperature of OH by two-line method

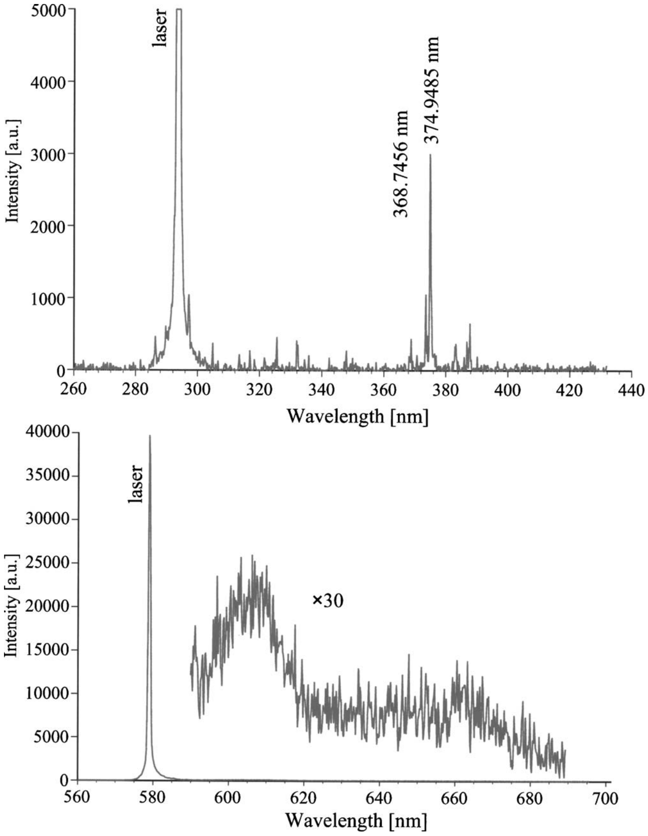


Fig. 5 LIF spectra: (*a*) Fe and (*b*) FeO

$$\frac{\beta_{\text{theo}}^2}{(1-\beta_{\text{theo}})N_{m,o}} = K(T) \tag{3}$$

where $N_{m,o}$ is the total number density of the iron-containing species introduced into the flame by spraying. The equilibrium constant, $K(T)$, is given by

$$K(T) = Z_{\text{Fe}}(T)Z_{\text{O}}(T)/Z_{\text{FeO}}(T) \tag{4}$$

The overall partition functions $Z(T)$ are expressed as the product of the partition function of the system involved, i.e.,

$$Z_{\text{Fe}}(T) = Z_{\text{tr,Fe}}(T)Z_{e,\text{Fe}}(T), \quad Z_{\text{O}}(T) = Z_{\text{tr,O}}(T)Z_{e,\text{O}}(T) \tag{5}$$

$$Z_{\text{FeO}}(T) = Z_{\text{tr,FeO}}(T)Z_{\text{vib,Fe}}(T)Z_{\text{rot,Fe}}(T)Z_{e,\text{FeO}}(T) \tag{6}$$

where the subscripts tr, vib, rot, and e stands for translational, vibrational, rotational and electronic systems.

Finally, the redox index, η_{rdx} can be obtained by taking a ratio of the experimentally and theoretically obtained degrees of atomization as follows:

$$\eta_{\text{rdx}} = \beta_{\text{exp}}/\beta_{\text{theo}} \tag{7}$$

If the combustion field is neutral in terms of the redox atmosphere, the degree of atomization is totally controlled by the simple thermal dissociation. Consequently, $\beta_{\text{exp}}=\beta_{\text{theo}}$ and $\eta_{\text{rdx}}=1$. The redox index, η_{rdx} becomes higher if the atmosphere during combustion is reductive, since the degree of atomization becomes higher because of the more atoms produced through reducing reactions than those expected by the simple thermal dissociation. In contrast, η_{rdx} becomes smaller than unity for the oxidative atmosphere.

The rotational temperature profiles were obtained with the conventional two-line method [3] by simultaneously measuring the LIF profiles of a pair of rotational lines, $R_2(5)$ and $Q_1(10)$, of OH. Assuming that the ratio of the quenching effect term for the two lines is independent of the temperature, the ratio of the former to the latter intensity, R_{OH} gives the rotational temperature, on the basis of the following equation:

$$R_{\text{OH}} = \frac{I_1 B_1 (2J_1'' + 1)}{I_2 B_2 (2J_2'' + 1)} \exp\left\{\frac{(E_2 - E_1)}{kT}\right\} \tag{8}$$

where I is the spectral intensity of the laser radiation, the subscripts 1 and 2 stand for the two lines, and J'', B, and E are the rotational quantum number, the transition probability for absorption and the energy, of the lower level to be excited by the laser radiation, respectively. The calibrations were made with an R-type thermocouple at several points in the flame.

The wavelengths and partition functions used in this work are summarized in Table 1. The atomic LIF of Fe is based on the three-level scheme as shown in Fig. 1.

Experimental

Figure 2 shows a schematic diagram of the experimental apparatus used in the study. The system mainly consists of the two parts: (i) the two-color PLIF system (Tokyo Instrument Co.) and (ii) the premixed propane-air planar flame supported on the burner. The instrumental specifications and the measurement conditions employed in this study are given in Table 2.

The two-color PLIF system employed in this study consists of two conventional lasers that are tunable independently. This system is used to simultaneously acquire the profiles of Fe and FeO profiles and also to obtain the profiles of OH rotational temperature by the two-line method. The dye lasers are excited with the

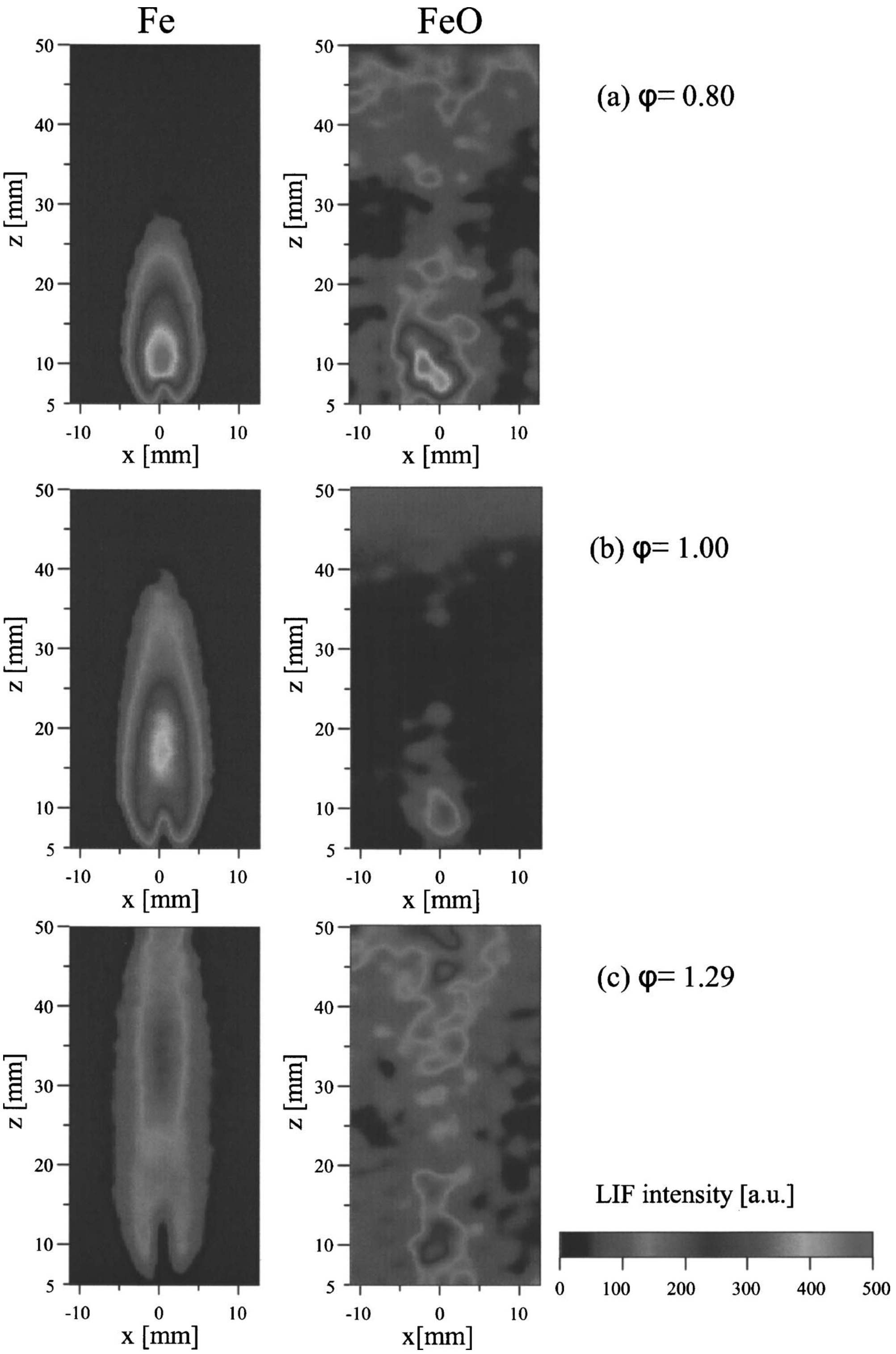


Fig. 6 Profile of LIF intensity

overtone radiation at 503.1 nm, emitted by the Nd-YAG (yttrium aluminium garnet) lasers. The radiation emitted from the dye lasers is frequency doubled and combined in a beam. Subsequently, it is formed into a plane sheet through the use of a cylindrical lens, which irradiates the analytical section of the flame in the transverse direction. The fundamental frequency was exceptionally used to excite the FeO at the longer wavelength of 579.10 nm. The two-color lasers are operated at a rate of 10 pulses/s with pulse duration of 8 ns and a delay of several nanoseconds between the first and second pulse. The two ICCD cameras are fitted with an optical band pass filter to remove the laser stray radiation and a boxcar integrator synchronously to acquire and accumulate the resulting LIF profiles. The trigger pulses are fed to the lasers and the ICCD cameras by the pulse generator. The preceding pulse delay of several nanoseconds from the trigger pulse is also applied for excluding the undesirable scattering signals and the electric noise. The overall process is computer controlled. Two-dimensional corrections were made for the nonuniformity of

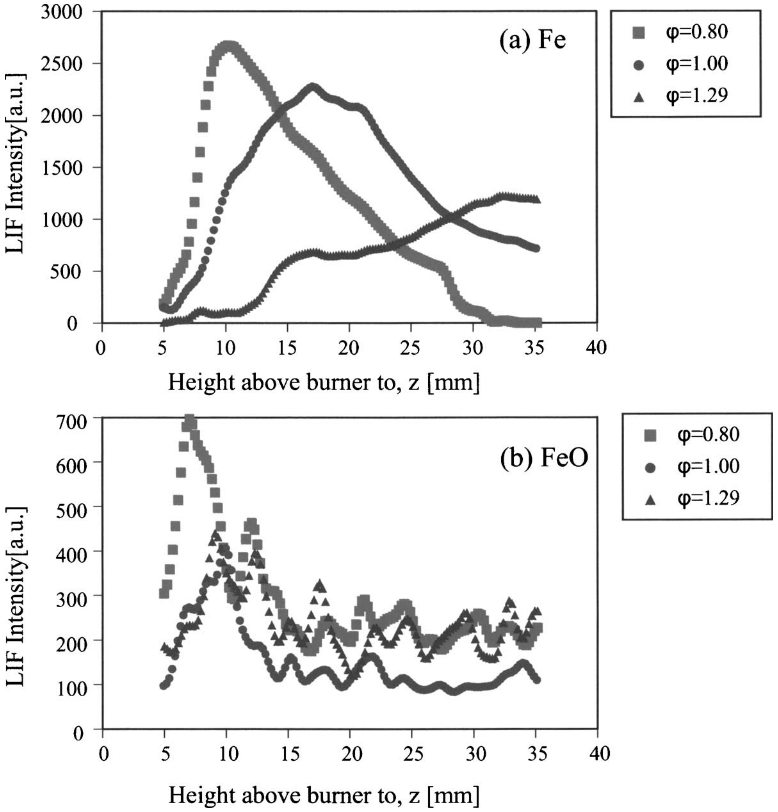


Fig. 7 LIF distributions of Fe and FeO above burner top

ICCD response and transmission efficiency of the filter using a uniform intensity source with flat intensity profile.

The burner used in this study is a conventional slot type having a nebulizer/mixing chamber for acetylene/air flame atomic absorption spectrometry. The slit width is widened (1.3 mm in width and 100 mm long) for propane/air combustion. The solution for chemical seeding is nebulized with air and mixed with the propane fuel and auxiliary air. The combustion conditions used are listed in Table 3.

By spaying a ferric nitrate ($Fe(NO_3)_3$) solution of 8.2 $\times 10^{-3}$ mol/l into the flame at an uptake rate of 2.0 ml/min, the LIF profiles of Fe and FeO were acquired for 20 s, or 200 laser pulses by the ICCD cameras. A computer to reconstruct the profiles of the degree of atomization and the redox index processed the time-averaged images obtained. The solution was prepared by dissolving a reagent of analytical grade in pure water of micropore filtration grade just prior to use.

Results and Discussion

Figure 3 shows the LIF profiles of the rotational lines of $R_2(5)$ (left column) and $Q_1(10)$ (right column) at three different equivalence ratios. These profiles show the relative number density of OH molecules on the lower levels, or approximately that of OH molecules as a cross-sectional view in the transverse direction. The results show that the number density of OH molecules is more abundant in the outer regions on both sides than in the middle region, including the plane perpendicular to the burner slit. This is mainly due to the small linear velocity in the outer region and to the diffusion of the ambient air. The number density decreases under fuel-rich conditions because the combustion rate is lowered and, consequently, the flame becomes longer. Figure 4 shows the profiles of the OH rotational temperature at three different equivalence ratios obtained on the basis of Eq. (8) and with the LIF profiles shown in Fig. 3. It is more clearly demonstrated by these figures that the flame volume (the length and thickness) becomes larger with the increase in equivalence ratio. The higher temperature is localized in the outer and upstream regions on both sides. It is well known that in the case of premixed flames, the ignition is initiated in the upstream region close to the burner exit because the mixing process is not the controlling factor in the case of diffusion flames. Consequently, the higher temperature is localized in the upstream region.

Figure 5 shows the atomic LIF spectrum (upper curve) of Fe and molecular LIF of FeO (lower curve). As indicated in the three-level diagram in Fig. 1, the fluorescent lines of Fe appear at the wavelengths of 368.7456 and 374.9485 nm. The FeO fluorescence spectrum shows mainly the (0,0) vibrational transition at the wavelength of 605.2 nm with the excitation at the (1,0) vibrational head at 579.4 nm.

Figure 6 shows the LIF profiles of Fe (left column) and FeO (right column) for three different equivalence ratios of 0.8, 1.0, and 1.29. The profiles indicate the relative number densities of Fe and FeO on the ground states, respectively. Since the flame temperature is relatively low for the excitation level energies of Fe atoms and FeO molecules, the profiles indicate, approximately, the relative number density of these species, respectively. The number density of Fe is higher in the middle region including the central plane above the burner slit where the temperature is lower (see the left photos in Fig. 6). If the simple thermal process, $FeO \rightarrow Fe + O$, prevails in the combustion field, the number density must be lower in this inner region and higher in the outer regions. The contrary experimental result suggests that the atomization process is not controlled by the simple thermal dissociation but by certain reactions with some intermediate reducing species, such as CH, C_2, H_2, hydrocarbon radicals, etc., that exist in the flame. Only with the LIF of Fe atoms, however, can the net redox atmosphere not be estimated as exemplified by the difference in the Fe profiles. The reducing power is expected to increase in the flame at higher equivalence ratios. The result shows the dependence to the contrary, or the increase and decrease in the number

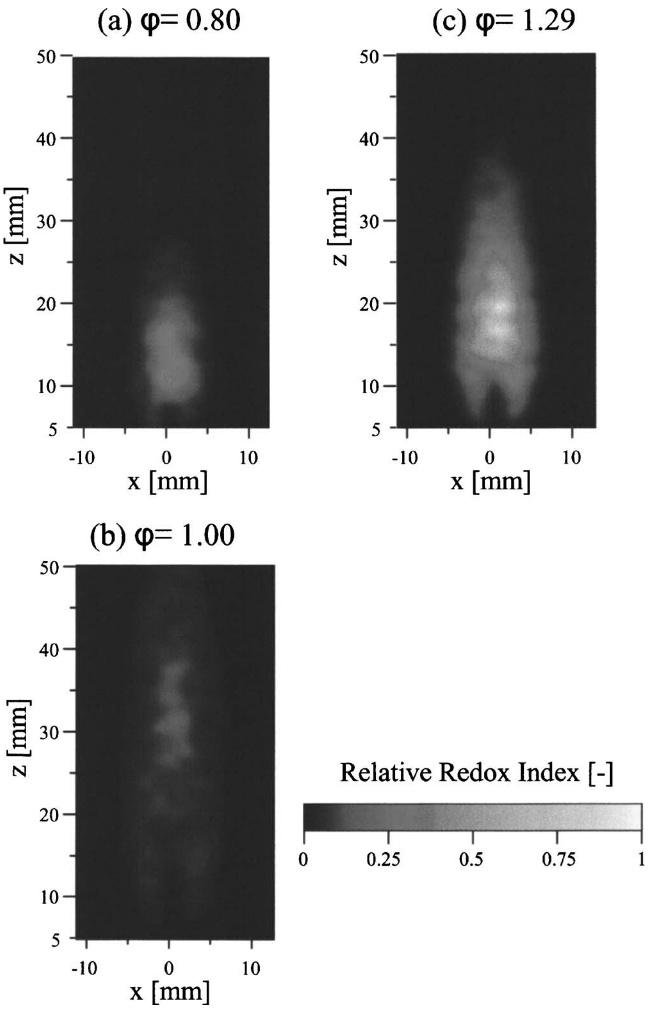


Fig. 8 Profile of relative redox index

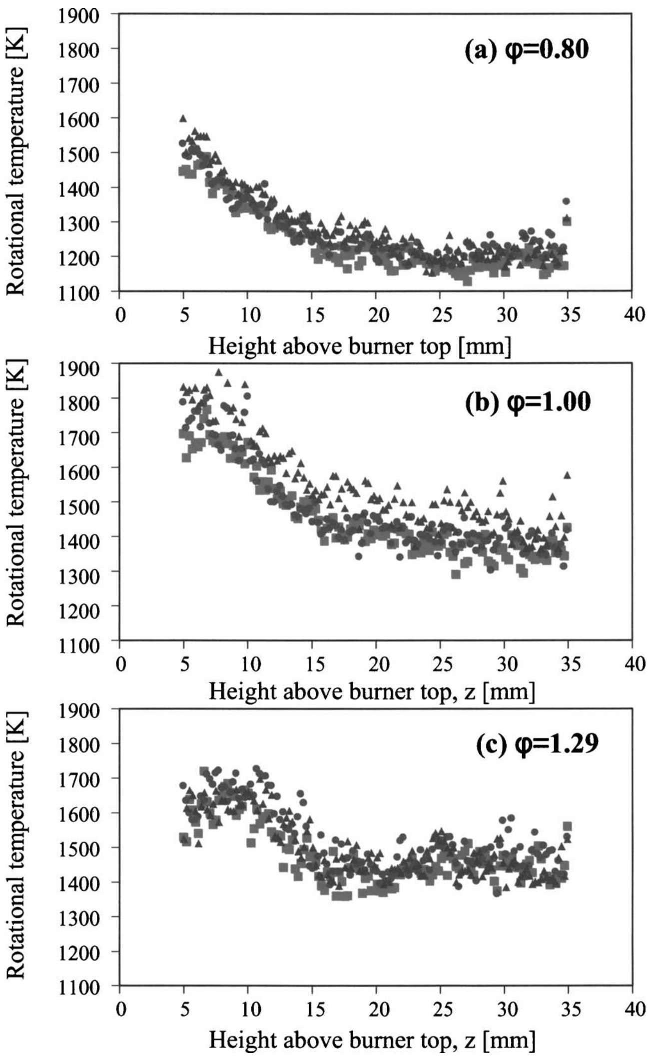


Fig. 9 Effect of spray on rotational temperature: ▲ without spray ■ water spray, ● spray of $Fe(NO_3)_3$ solution

density of Fe atoms (compare the left three photos in Fig. 6). This is mainly because of the increase in flame volume, leading to the dilution of Fe vapor. This can also be seen form the right three photographs of FeO shown in Fig. 6. Figure 7 shows the LIF distribution along the central axis of the flame and supports the above-indicated tendency. One can note that there are some fluctuations to the FeO LIF intensity in the results. This is possibly due to the low-intensity signal from FeO fluorescence (as also shown in the spatial distribution results in Fig. 6).

Figure 8 shows the profiles of relative redox index obtained by defining the maximum value as unity, or $\eta_{\mathrm{rdx}}/\eta_{\mathrm{rdx,max}}$. The results show that a moderate reductive region (reductive index of 0.5) expands in both axial and radial directions (compare Figs. 8(*a*) and 8(*b*)). A highly reductive index of around unity expanded over a much wider area is achieved, as shown in Fig. 8(*c*). Thus, under a high-equivalence-ratio condition, one can clearly quantify the reductive atmosphere using the redox index presented in this study. The burner we used here is a premixed mixture diffusing into surrounding ambient air. The increase in the redox index in the central region of the flame is due to the local fuel-rich conditions encountered in the flame and insufficient diffusion of the oxidant from the surroundings.

Figure 9 shows the influence of chemical seeding on the temperature profile above the burner. The water lowers the temperature (compare the solid triangles and rectangles in Fig. 9) This is due to the latent heat of water vaporization. The introduction of the ferric nitrate increases the temperature slightly, particularly at higher equivalence ratios (compare the solid rectangles and circles). This is probably attributable to the thermal dissociation of nitrate to produce oxygen species, i.e., $NO_3^- \rightarrow NO + 2O$ or O_2, leading to the enhancement of the combustion reaction.

Conclusions

A new method, *chemical seeding/laser-induced fluorescence (CS/LIF)*, has been developed and successfully applied to quantitatively determine the reducing/oxidizing atmosphere during combustion. The chemical seeding element, Fe, is nontoxic and easily available. In industrial furnaces used for the iron/steel process, iron oxides exist without any seeding. This method can be used to determine the profiles of the redox index in furnaces where the product quality is much affected by the local oxidative or reducing conditions in the furnace flame.

The difficulty of applying LIF systems to practical furnaces in severe environments still exists. However as compared to the general chemical seeding/spontaneous emission spectrometry (CS/SES), this method is safe, reliable, and very practical. The former method had the disadvantage of providing results with nonreal spatial distribution. Compact and rug LIF systems are the key for the application of laser diagnostics to industrial combustion systems.

References

[1] Kubota, M., Kitagawa, K., Arai, N., and Gupta, A. K., 2002, "Profiling of Redox Index During Combustion to Monitor C/C Composite Degradation in Flames," J. Propul. Power, **18**(2), pp. 372–375.
[2] Kitagawa, K., Ide, Y., and Takeuchi, T., 1980, "Spectroscopic Determination of Atomization Efficiency (CuCl→Cu+Cl) in an Air-Acetylene Flame," Anal. Chim. Acta, Vol. **113**, pp. 21–32.
[3] Cattolica, R., 1981, "OH Rotational Temperature From Two-Line Laser Excited Florescence," Appl. Opt., **20**, pp. 1156–1166.
[4] De Galan, L., Smith, R., and Winefordner, J. D., 1976, "The Electric Partition Functions of Atoms and Ions Between 1500°K and 7000°K," Spectrochim. Acta, Part B, **23B**, pp. 521–525.
[5] Herzberg, G., 1950, *Molecular Spectra and Molecular Structure I. Spectra of Diatomic Molecules*, VanNostrand Reinhold, Princeton.
[6] Moore, W. J., 1956, *Physical Chemistry*, 2nd ed., Lowe & Brydone, London.
[7] Corliss, C. H., and Bozmann, W. R., 1994, *NBS Monograph 53*, U.S. Government Printing Office, Washington, DC.

H. S. Tan
Republic of Singapore Air Force,
Air Logistics Department,
Propulsion Branch,
303 Gombak Drive,
01-44 Singapore 669645

Fourier Neural Networks and Generalized Single Hidden Layer Networks in Aircraft Engine Fault Diagnostics

The conventional approach to neural network-based aircraft engine fault diagnostics has been mainly via multilayer feed-forward systems with sigmoidal hidden neurons trained by back propagation as well as radial basis function networks. In this paper, we explore two novel approaches to the fault-classification problem using (i) Fourier neural networks, which synthesizes the approximation capability of multidimensional Fourier transforms and gradient-descent learning, and (ii) a class of generalized single hidden layer networks (GSLN), which self-structures via Gram-Schmidt orthonormalization. Using a simulation program for the F404 engine, we generate steady-state engine parameters corresponding to a set of combined two-module deficiencies and require various neural networks to classify the multiple faults. We show that, compared to the conventional network architecture, the Fourier neural network exhibits stronger noise robustness and the GSLNs converge at a much superior speed. [DOI: 10.1115/1.2179465]

1 Introduction

The application of neural networks to aircraft engine fault diagnosis has been mostly constrained to multilayer feed-forward networks with sigmoidal transfer functions operating on the basis of back-propagation learning (BPFN) [1–4] and radial basis function (RBF) networks [5]. Recently, there have been works that explore alternative methods of neural computation applied to fault diagnostics. Prominently, in [6], a genetic algorithm (GA) was first attempted on the fault analysis of advanced cycle gas turbines. Also in [7], a GA-based gas path fault diagnosis of a gas turbofan engine was also presented. Venturing beyond the BPFN platform and back-propagation learning can bring us new surprises as much as it deepens our understanding of the virtues and limitations of BPFN in the area of fault diagnostics. For instance, in [8], the use of Kalman filters, which are essentially linear model-based approximators, was shown to offer a slight advantage over conventional back-propagation networks in isolating single-engine modular faults. And in [9], a comparative study was performed among different neural network architectures—multilayer perceptrons, linear filters, self-organizing maps, and learning vector quantization (LVQ) networks—in diagnosing rotating machinery faults.

It is in this spirit that we propose two novel approaches to aircraft engine fault diagnostics here—first, a Fourier neural network, which synthesizes the approximation capability of multidimensional Fourier transforms and gradient-descent back propagation; and second, a class of generalized single hidden layer networks (GSLN), which self-structures at each iteration via the well-known Gram-Schmidt orthonormalization procedure. Both network methodologies are theoretically well founded but have never surfaced previously in literature as artificial intelligence applications to engine fault diagnostics. We will study their working principles and applicability in a fault diagnostics problem, as compared to two-layer BPFN models that have sigmoidal hidden layers.

We are interested in the generalization performances of the different neural networks in diagnosing multiple modular faults of the F404 engine. Deterioration of each engine module (i.e., reduction in each module's efficiency) induces changes to aerothermodynamic gas path properties, and thus, the deviations of an engine's parameters (like exit temperatures) from those of a presumedly nominal one are natural signatures of modular problems. These residuals are thus appropriate inputs into any fault diagnostics system. In this paper, we employ the GE digital simulation program for F404 engine—*Cycledeck* [10]—to supply engine data corresponding to two-module faults, the modules being the high-pressure compressor (HPC) and turbine (HPT), low-pressure compressor (LPC) and turbine (LPT). It turns out that for single fault diagnostics, the problem is actually linearly separable (see Fig. 1); each module generates a distinct line (2–5% deficiency) in every two-parameter phase space. All the neural networks of interest yield perfect generalization accuracy, and, indeed, for such a case, it is well known (and shown in [11]) that a single-layer perceptron model is sufficient. On the other hand, the situation complexifies when extended to multiple modular faults. The data distribution reveals overlapping regions in the two-parameter phase spaces corresponding to different two-module faults (see Fig. 1). This necessitates use of more than two thermodynamic variables to distinguish between fault cases, and, indeed, as later shown, this is the case.

The performance variables we used are: (i) WFE, engine fuel flow rate, (ii) N2, high-pressure rotor rotational speed; (iii) T5H, harness total temperature; (iv) PT56, hot stream total pressure before mixing; and (v) PS3, high-pressure compressor exit static pressure. The choice of these parameters is largely motivated by their measurability in test cells and the F404 Engine Health Monitoring System (EHMS) [12]. Although the neural networks developed in this paper are trained using simulated data, when enough physical data have been accumulated in the future, moving on to a real-life application will be a straightforward process. The data were all generated at standard day conditions, but there is no true loss of generality here because correction factors for engine parameters due to ambient temperatures and pressures can be derived following the procedure of [13].

The paper is organized as follows: we begin by introducing the mathematical background behind Fourier neural networks and GSLNs, outlining their operating principles and algorithms. We

---

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received October 4, 2004; final manuscript received October 17, 2005. Review conducted by W. Rhoden.

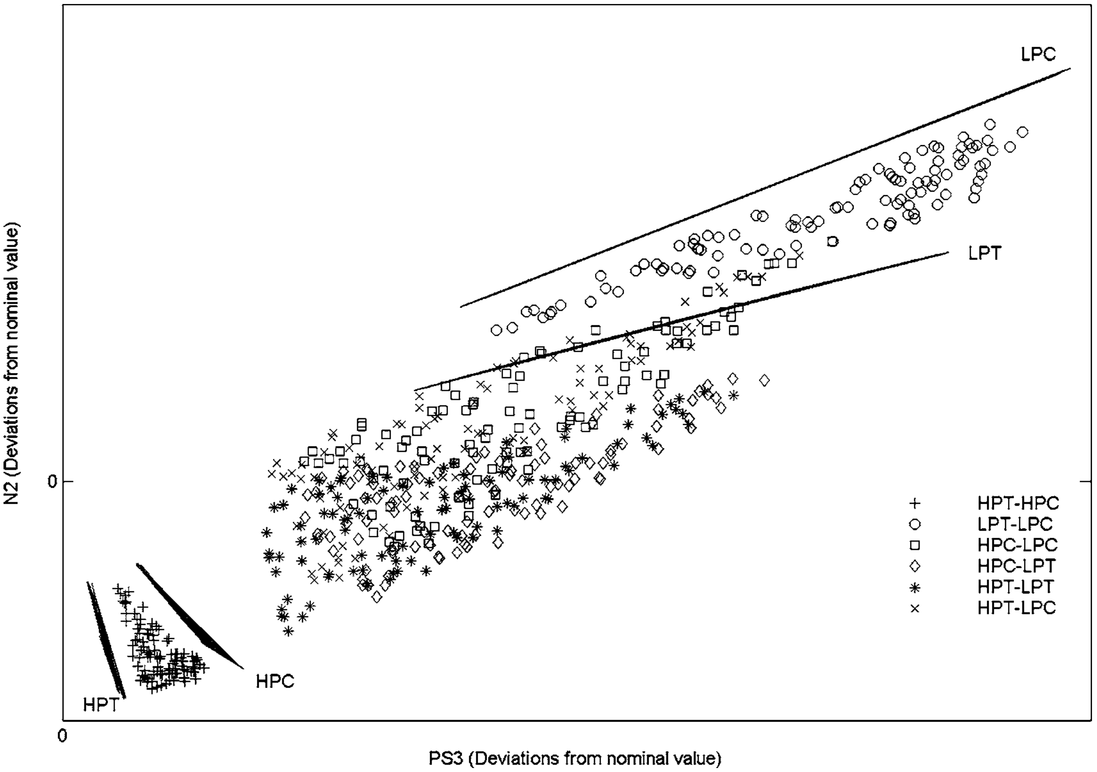


Fig. 1 Data distribution of two module faults

then discuss, in detail, how we formulate the multiple fault diagnostics problem in the framework of neural networks. This includes a careful discussion on our choice of output representation and our decision rules. Various experimental results are presented next, including a comparison of the networks' noise robustness and convergence speeds. The paper ends with conclusions regarding the novel neural networks and some suggestions for future work.

2 Theoretical Background

2.1 Fourier Neural Networks. Consider a mapping f that takes a D-dimensional input vector $\vec{x}$ to a scalar output; hence, $f:\Re^D \rightarrow \Re$. A conventional L-layer feed-forward neural network with a scalar output is a function approximator of the form

$$f(\vec{x}) = \varphi_L\left(\sum_k W_{1k}\varphi_{L-1}\left\{\sum_j V_{kj}\varphi\left[\cdots\varphi_1\left(\sum_i U_{li}x_i\right)\right]\right\}\right) \qquad (1)$$

with $\vec{x} \in \Re^D$, W, U, V being the layers' weights and the transfer functions φ_i being sigmoidal or linear functions. If the weights are trained via back-propagation algorithm, then the nested sigmoidal scheme of Eq. (1) is a *universal* approximator [14]. To be more specific, it was shown in [15] that a continuous function can be arbitrarily well approximated by a BPFN with only a single hidden layer, where each unit in the hidden layer has a continuous sigmoidal nonlinearity, and it was further revealed in [16] that a BPFN with two hidden layers provides a converging approximation for any nonlinear mapping. Now, it turns out that there is another model of neural computation [17] for which the network output assumes the following form:

$$f(\vec{x}) = \int_{D_2} C(\vec{k})\varphi(\vec{x},\vec{k})d\vec{k} \qquad (2)$$

Observe that in Eq. (2), the transfer functions φ_i are functions of the input $\vec{x}$ and another variable $\vec{k} \in \Re^{D_2}$, weighted by the coefficient C and integrated with respect to $\vec{k}$. Note that, in general, the dimension of $\vec{k}$ needs not be the same as the input dimensionality. In the language of functional analysis [18], Eq. (2) represents an integral operator mapping the input space to the output space with kernel $\varphi(\vec{x},\vec{k})$. Furthermore, if $\varphi(\vec{x},\vec{k}) = e^{i\vec{x}.\vec{k}}$, then up to a multiplying constant, C is the Fourier transform of f on $\vec{k}$.

For computational purpose, we consider the discretized version of Eq. (2) as

$$f(\vec{x}) = \sum_{k_1 \cdots k_D} C(k_1 \cdots k_D)\exp\left(i\sum_j x_j k_j\right) \qquad (3)$$

where we have taken the Fourier kernel and, thus, let $D_2 = D$. When $\vec{x}$ and $\vec{k}$ are discretized from 1 to an integral index N, we arrive at the discrete Fourier transform. Equation (3) can be interpreted as having a network structure in Fig. 2, where we have let $\vec{W}$ denote $\vec{k}$, and the subscript i in C_i being a multiindex $\{k_1, k_2, \ldots, k_D\}$.

This is a single hidden layer neural network with the complex exponential transfer function and complex-valued weights for the outer layer. By Fourier theory, the weights $C(k_1, \ldots, k_D)$ in Eq. (3) are obtained via inversion

$$C(\vec{k}) = \sum_{x_1 \cdots x_D} f(\vec{x})\exp\left(-i\sum_j x_j k_j\right) \qquad (4)$$

Instead of Eq. (4), we can obtain $C(\vec{k})$ via gradient descent methods since we have interpreted (3) as having a neural network structure. Consider a two-dimensional problem with a real scalar output f. Letting $C(\vec{k}) = C_R(\vec{k}) - iC_{\mathrm{Im}}(\vec{k})$, we have from Eq. (3)

$$f(x_1,x_2) = \sum_{k_1=1,k_2=1}^{N,N} \left\{ \begin{array}{l} C_R(k_1,k_2)\cos\left[\left(\dfrac{2\pi(k_1-1)x_1}{N}\right)+\left(\dfrac{2\pi(k_2-1)x_2}{N}\right)\right] \\ +\, C_{\mathrm{Im}}(k_1,k_2)\sin\left[\left(\dfrac{2\pi(k_1-1)x_1}{N}\right)+\left(\dfrac{2\pi(k_2-1)x_2}{N}\right)\right] \end{array} \right\}$$

$$= \underbrace{[C_R(1,1)C_R(1,2)\cdots C_R(N,N)C_{\mathrm{Im}}(1,1)\cdots C_{\mathrm{Im}}(N,N)]}_{\text{Weight vector of dimension } 2N^2} \left(\begin{array}{c} 1 \\ \cos\left(\dfrac{2\pi x_2}{N}\right) \\ \vdots \\ \cos\left(\dfrac{2\pi(N-1)x_2}{N}\right) \\ \cos\left(\dfrac{2\pi x_1}{N}\right) \\ \vdots \\ \cos\left(\dfrac{2\pi(N-1)(x_1+x_2)}{N}\right) \\ \sin\left(\dfrac{2\pi(N-1)(x_1+x_2)}{N}\right) \end{array} \right) \tag{5}$$

Equation (5) is the explicit expression of the Fourier neural network with the weight vector learned via gradient descent. A multiple output system can be constructed by simply adding more neurons (and, thus, weight vectors). Furthermore, we can constrain the network's output to a bounded interval by passing the output (3) through another transfer function. For example, if we require the network output to fall within $[-1,1]$, the final model can then be

$$f(x_1,x_2) = 2\left[1+\exp\left(\sum_{k_1=1,k_2=1}^{N,N} C(k_1,k_2)\exp[-i(k_1x_1+k_2x_2)]\right)\right]^{-1} - 1 \tag{6}$$

Next, we remark on the convergence properties of Eq. (6) following [17]. For a multidimensional Fourier series, we have the following theorem [19]:

THEOREM 1. *If* $f:\vec{X}\to\vec{Y}$ *is continuous with period* 2π *in every subset of components, then the Fourier series for f converges uniformly to f.*

It is then straightforward to show that Eq. (6) can approximate arbitrarily well every continuous function defined on $[-1,1]^D$. As pointed out in [17], the difference between the Fourier neural network formalism and discrete Fourier transform lies in the fact that statistical accuracy is recovered in the latter while the former can converge via back-propagation learning to a near exact accuracy. The Fourier neural network effectively embeds gradient descent-based back-propagation learning within the structure of a finite Fourier series. It is natural to extend the Fourier neural network to other models of the form Eq. (2). For example, the coefficients of the wavelet transform can be recovered via back-propagation learning instead of the usual inversion formula. Indeed, *wavelet* neural networks [20] can be constructed similarly and have been used in many applications that involve dynamic time behavior [21].

In our experiments, we will investigate the generalization capabilities of Eq. (6) using the optimal integral parameter N for different input dimensionalities. The value of N controls the number of adjustable weights, and, thus the degrees of freedom of the approximation series.

2.2 Generalized Single-Layer Networks. Consider an approximation of an input-output relationship of a form similar to Eq. (3) but with the complex exponential functions replaced by generalized basis functions ϕ

$$f(\vec{x}) = \sum_{a=1}^{N} W_a\phi_a(\vec{x}) \tag{7}$$

where $\vec{x}$ is the input, $\{W_a\}$ is the set of weights to be adjusted later on and $\{\phi_a\}$ is a set of N nonlinear basis functions. Given K training input-output pairs $\{\vec{x}_\mu, Y_\mu\}$, $\mu=1,2,\ldots,K$, we define the network cost function E as the sum of mean-squared error as usual

$$E = \frac{1}{K}\sum_{\mu=1}^{K}[f(\vec{x}_\mu)-Y_\mu]^2 \tag{8}$$

Minimizing Eq. (8) with respect to the weights yields the equations

$$\sum_{m=1}^{N} W_m\left(\sum_{\mu=1}^{K}\phi_m(\vec{x}_\mu)\phi_i(\vec{x}_\mu)\right) = \sum_{\mu=1}^{K} Y_\mu\phi_i(\vec{x}_\mu) \tag{9}$$

Given a set of distinct $\vec{x}_\mu$, we have a regression solution that is a global minimum in contrast to the local minimum that is possibly encountered in back-propagation-based methods. Equation (9) can be solved quickly via Gram-Schmidt orthonormalization [22]. Analogous to the geometric vector space, we begin by defining the inner product of two basis functions as $\langle\phi_a(\vec{x})|\phi_b(\vec{x})\rangle \equiv \Sigma_{\mu=1}^{k}\phi_a(\vec{x}_\mu)\phi_b(\vec{x}_\mu)$, and the norm $\|\phi_a(\vec{x}\| = \sqrt{\langle\phi_a|\phi_a\rangle}$. Two functions are orthogonal to each other if their inner product is zero. We proceed to construct a set of orthogonal functions in place of the original $\{\phi_a\}$ by the following definitions:

$$\psi_1(\vec{x}) = \frac{\phi_1(\vec{x})}{\|\phi_1(\vec{x})\|},$$

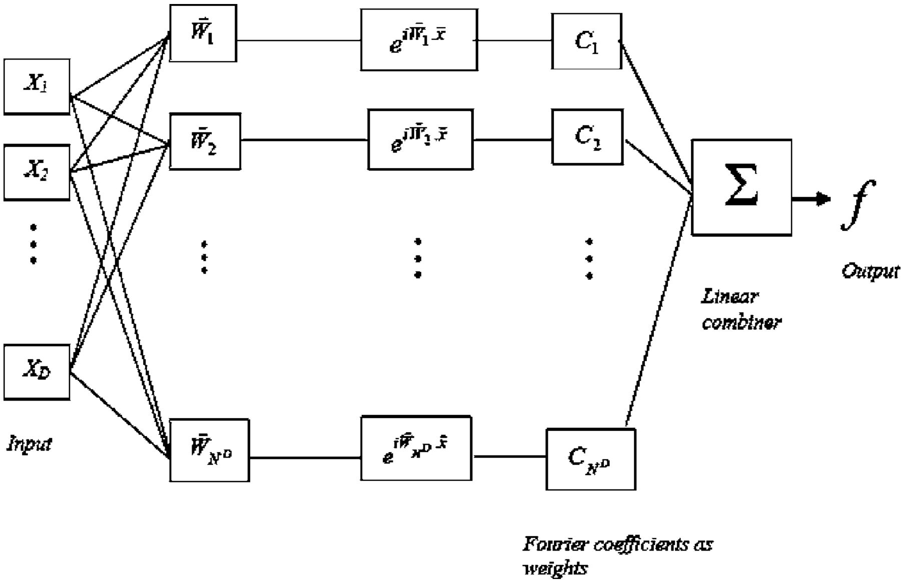


Fig. 2 The structure of Fourier neural network

$$\psi_l = \frac{\left(\phi_l(\vec{x}) - \sum_{i=1}^{l-1} \psi_i(\vec{x})\langle\psi_i(\vec{x})|\phi_l(\vec{x})\rangle\right)}{\left\|\left(\phi_l(\vec{x}) - \sum_{i=1}^{l-1} \psi_i(\vec{x})\langle\psi_i(\vec{x})|\phi_l(\vec{x})\rangle\right)\right\|}, \quad \forall\, l = 2,3,\ldots,N \tag{10}$$

This is the well-known Gram-Schmidt orthonormalisation, which yields

$$\langle\psi_i|\psi_j\rangle = \delta_{ij} \tag{11}$$

where δ_{ij} is the Kronecker δ function. From Eq. (10), $\{\psi_a\}$ is some linear transformation of $\{\phi_a\}$ and we can write the network output as

$$f(\vec{x}) = \sum_{a=1}^{N} \widetilde{W}_a \psi_a(\vec{x}) \tag{12}$$

where $\{\widetilde{W}_a\}$ is the new set of weights with respect to the orthonormalized basis functions. Equation (12) (or equivalently Eq. (7)) has the network structure depicted in Fig. 3.

Since $\{\psi_a\}$ is an orthogonal set, we can substitute Eq. (11) into Eq. (9) and this yields the solutions

$$\widetilde{W}_a = \sum_{\mu=1}^{K} Y_\mu \psi_a(\vec{x}_\mu) \tag{13}$$

The globally optimal network can thus be obtained easily. Essentially, just as in a RBF network, the input vector undergoes a nonlinear transformation and is mapped to a space of dimension N at the hidden layer. The basis functions are orthogonal in the normed space, and this lends an effective computation of the optimal weights. Geometrically, as pointed out in [23], the basis function of each added neuron spans the space orthogonal to the existing set of functions. If we denote the network error at epoch α by $E_\alpha(\vec{x})$, then at the end of the $N+1$ iteration, we have

$$E_N(\vec{x}) = E_{N+1}(\vec{x}) + \widetilde{W}_{N+1}\psi_{N+1}(\vec{x}) \tag{14}$$

Seeking the minimum $E_{N+1}(\vec{x})$ implies finding $\psi_{N+1}(\vec{x})$ most "aligned" with $E_N(\vec{x})$—this interpretation lends a geometrical meaning to the orthogonalization process. The learning algorithm is as follows [24]: we begin with a single hidden neuron and add neurons (basis functions) one at a time iteratively to the network, orthogonalizing them via Eq. (10) at each step. The iteration ends

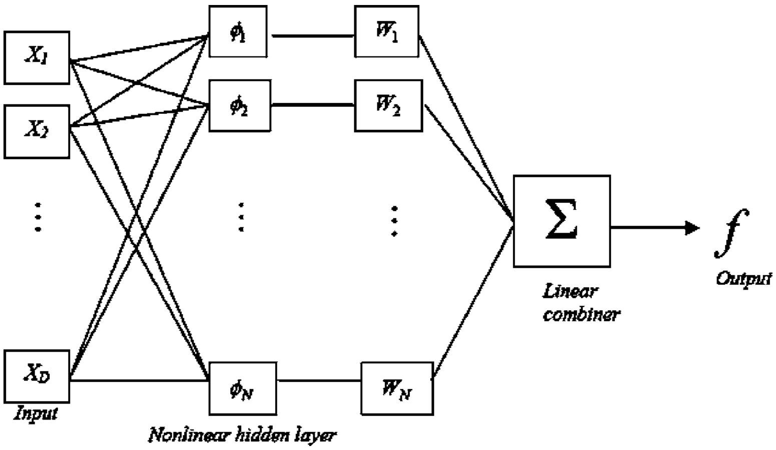


Fig. 3 The structure of GSLN

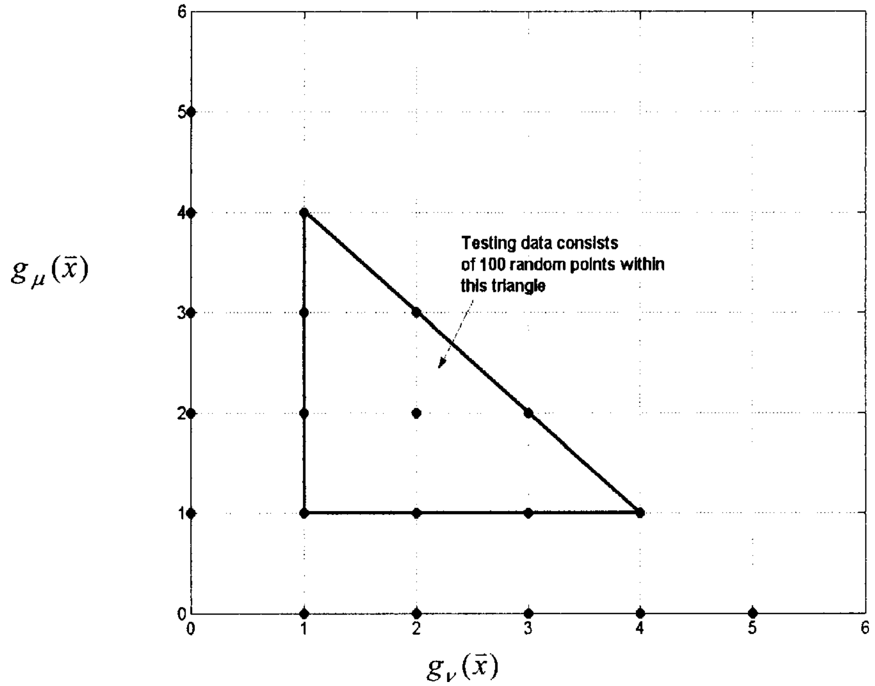


Fig. 4 Distribution of training and testing data

when the training error in Eq. (8) decreases to below a certain desired level.

The parameters that enumerate each basis function can be chosen randomly from bounded intervals. In this paper, we use the following two families of functions:

$$\text{Gaussian:} \quad G(\vec{\alpha},\vec{\beta},\vec{x}) = \exp\left(-\frac{\|\vec{x}-\vec{\alpha}\|^2}{2\beta^2}\right) \qquad (15a)$$

$$\text{Sigmoidal:} \quad S(\vec{a},\vec{b},\vec{x}) = \prod_{i=1}^{D}\{1+\exp[-a_i(x_i-b_i)]\} \qquad (15b)$$

where the function parameters $\vec{\alpha}$, $\vec{a}$, $\vec{b}$ are uniformly distributed random numbers drawn from the interval $[-1,1]^D$, and β in the interval $[0.5,2.5]$. We can also choose other types of basis functions of other symmetry features, see [25] for a comprehensive review. Also, the relationship between the Fourier neural network and GSLN is clear. Equation (12) takes the form of Eq. (3) with a being the multi-index $\{k_1,\ldots,k_D\}$ and complex exponential functions as basis functions. In the case of Fourier neural networks, the network does not grow with each iteration and the trigonometric functions are not orthogonal to each other (by our definition of inner product). Thus, the optimal weights of the Fourier networks cannot be obtained via Eq. (13).

3 Network Modeling

3.1 Data Preprocessing. We are interested in the generalization capabilities of various neural networks in mapping five engine performance variables—WFE, N2, T5H, PS3, and PT56—to all of the six possible two-module fault classes. The four modules of the F404 engine are HPT, HPC, LPT, and LPC. The input variables to the networks are simply the deviations of the various parameters from their nominal values. Besides the Fourier neural network in which the input values are normalized between 1 and the integral index N, the GSLNs and back-propagation-based feed-forward networks will be fed with input normalized within the interval $[-1,1]$. This follows from the well-known heuristics rule (see, for example, [14]) that each input variable should be preprocessed so that its mean value is close to zero.

We now come to the question of what constitutes an appropriate output representation for our classification problem. Since there are six multifault classes, a straightforward choice would be to have a six-dimensional (6D) output vector $F_i(\vec{x}), 1 \leq i \leq 6$ of which each component denotes the degree to which the input $\vec{x}$ belongs to each fault class. However, in this case, we have the prior knowledge that these fault classes are not physically independent from each other. Consider, for example, the fault classes HPT-HPC and HPT-LPT. The engine parameters corresponding to both fault scenarios share a common origin in the HPT deterioration. Thus, we can choose to have a four-dimensional (4D) output vector $F_i(\vec{x}), 1 \leq i \leq 4$ for which each component denotes instead the degree to which each module has deteriorated. A natural output decision rule is the Bayes rule stating: Classify input $\vec{x}$ as belonging to the multifault class of modules μ and ν if

$$F_\mu(\vec{x}), F_\nu(\vec{x}) > F_i(\vec{x}) \quad \text{for all} \quad i \neq \mu,\nu$$

The data sets are generated by a F404 engine simulation program, which yields precise estimates of the modular deficiencies in the range 0–5%. This leads us to consider the continuous mapping

$$F_i(\vec{x}) = 0.16 g_i(\vec{x}) + 0.1 \qquad (16)$$

where $g_i(\vec{x})$ is the ith module's deficiency in percentage. Equation (16) maps all output component values to the interval [0.1,0.9], where 0.1 represents the nominal state and 0.9 represents the maximum deficiency level in our experiments.

Now to be realistic, accurate numerical estimates of the modular deficiencies are difficult if not impossible to obtain, in particular, for test cell settings. If a neural network is designed for learning via real-life data, it must admit target values that represent at most fuzzy knowledge of the deficiency levels. For example, instead of Eq. (16), we could have simply

$$F_i(\vec{x}) = \begin{cases} \varepsilon & \text{if module is nominal} \\ 1-\varepsilon & \text{if module is faulty} \end{cases} \qquad (17)$$

where ε is a small positive constant. Now Eq. (17) can be made more sensitive to our rough knowledge of how serious each engine module has deteriorated by letting

Table 1 Networks' accuracies at different input dimensions, averaged over all parameter combinations

Input dimensions	Fourier (N=2)	Fourier (N=3)	Fourier (N=4)	Fourier (N=5)	BPFN (2 hidden neurons)	BPFN (10 hidden neurons)	Sigmoidal GSLN	Gaussian GSLN
2	43.5	47.58	46.85	41.7	44.98	51.78	50.53	50.02
3	71.47	67.75	52.22	37.72	57.78	73.73	74.13	74.23
4	90.90	78.10	41.67	31.77	64.14	92.6	97.37	95.33
5	99.83	65	31	29	66.67	99.87	100	100

$$F_i(\vec{x}) = \varepsilon + \mu(1 - 2\varepsilon) \tag{18}$$

where $0 \leqslant \mu \leqslant 1$ can be interpreted as a fuzzy membership value. Note that both Eqs. (17) and (18) keep $F_i(\vec{x})$ within the range $[\varepsilon, 1-\varepsilon]$. In our experiments, we adopt two different output representations: Eq. (16), which we call *continuous targets*, and Eq. (18) with $\varepsilon = 0.1$, and

$$\mu = 1.25(0.01g + 0.75)\Theta(g - 1) \tag{19}$$

with $\Theta(g-1)=1$ for all $g>1$ and 0 otherwise. The choice of membership function implies that for deterioration of $<1\%$, the module is described as nominal with $F_i(\vec{x})=0.1$, while for deterioration exceeding 1%, the membership function increases linearly with g, from 0.95 to 1, with $0.86 < F < 0.9$. The numeric constants in Eq. (19) are arbitrary but chosen so that we almost have Eq. (17). We call Eq. (19) *binary targets* due to its closeness to Eq. (17). Indeed, in real-life applications where the target output is not numerically quantifiable, binary targets become the obvious choice of representation.

Although we have utilized both Eqs. (16) and (19) in our experiments, results showed that the neural networks trained on the same input data for either of them did not produce significantly different results. Thus, we will only present results corresponding to binary targets for compactness, taking note that the various analysis and conclusions in Sec. 3.2 hold for continuous target choice as well. Finally, for each multifault class defined by modules μ and ν, our training data consists of all engine parameters simulated from all possible integer pairs of module deteriorations $\{g_\mu(\vec{x}), g_\nu(\vec{x})\}$ within the region $g_\mu(\vec{x})+g_\nu(\vec{x}) \leqslant 5$, which consist of 21 points in total (see Fig. 4). The testing data set consists of 100 random points in the domain $2 \leqslant g_\mu(\vec{x})+g_\nu(\vec{x}) \leqslant 5$, thus yielding a training to testing data size ratio of about 1:5.

3.2 Experimental Results. Table 1 shows the average accuracies in percentages of the various neural networks when presented the testing data sets. The back-propagation networks (BPFN) are two-layer networks each having a sigmoidal hidden layer and linear outer layer, with the numbers in the parenthesis in Table 1 indicating the number of neurons in the hidden layer. Both BPFN and Fourier networks have their weights learned via gradient-descent method with an adaptive learning rate [26] and momentum of 0.5, while the GSLNs have theirs updated through the procedure discussed in Sec. 2.2. The results were taken at 10,000 epochs for both BPFN and Fourier networks, and the optimal epoch (typically, in the range of 30–40 and beyond which the training error starts to increase) for the GSLNs. These results are obtained as averages over ten repetitions for each neural network.

Figure 5 displays the variation of the neural networks' performances with input dimensionality. Note that for BPFN, we have checked that increasing the number of hidden neurons beyond-ten did not improve our various results significantly. For each neural type, there exist network parameters such that perfect accuracy can be realized if all five engine variables were used as inputs. As we decreased the input dimensionality, the generalization performance shared a common decrease. This reminds us of Cover's theorem [27] that a complex pattern-classification problem cast in a high-dimensional space nonlinearly is more likely to be linearly

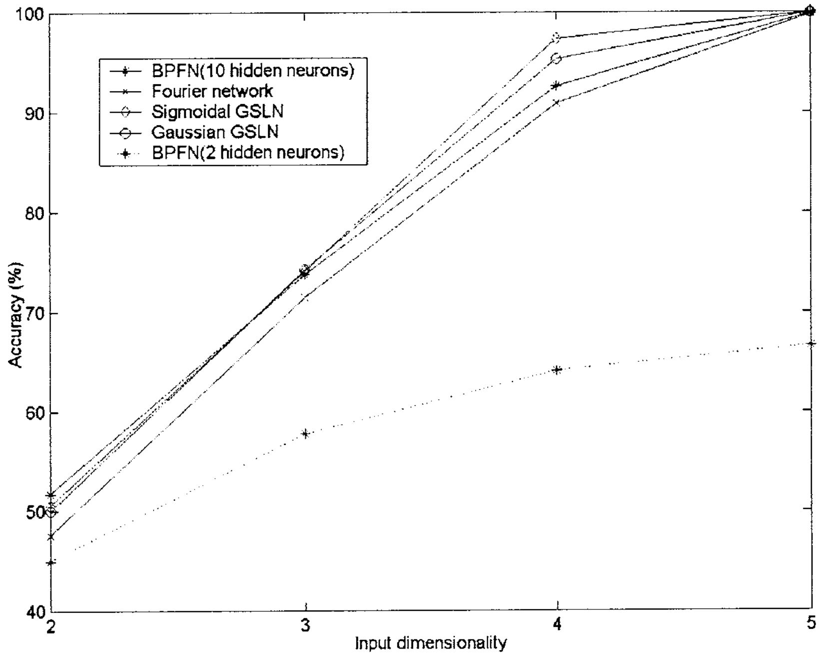


Fig. 5 Networks' accuracies at different input dimensions

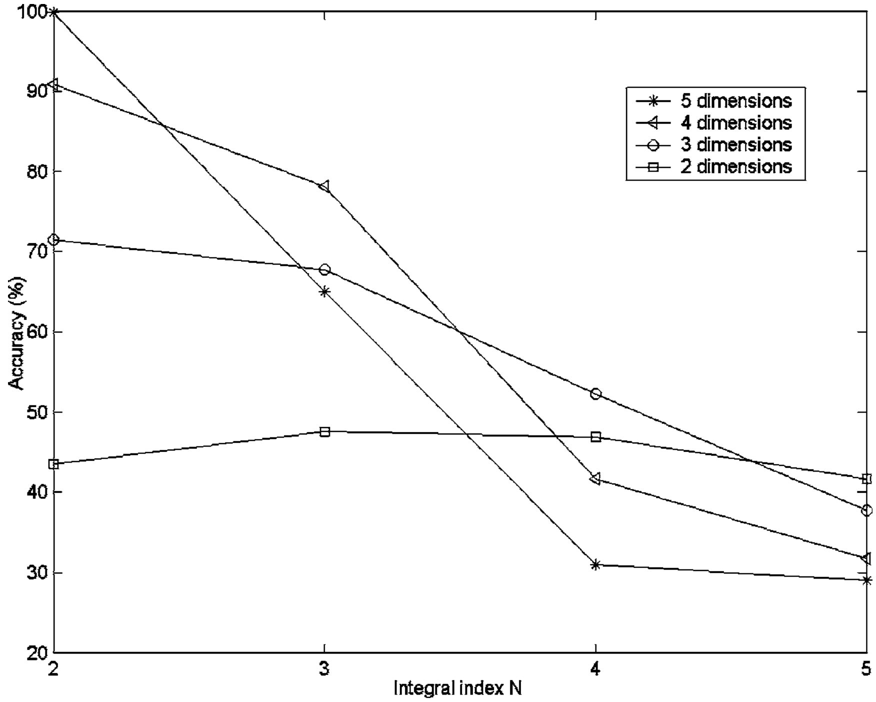


Fig. 6 Fourier neural networks' accuracies at different integral index *N*

separable than in a low-dimensional space. In this case, the abstract dimensions are actually physical input dimensions. In Fig. 6,we observe that the optimal value of N for the Fourier networks is 2 for three to five input dimensions and is 3 for two dimensions. This result is specific to our problem, but generally, increasing N beyond the optimum value leads to an overparametrization of the network model, which subsequently overfits. The nonlinear mappings between the input and output as provided by the neural networks can also be interpreted as four different hypersurface constructions—one for each component of the 4D output vector. For visualization purposes, we consider a two-dimensional input problem with $T5H$ and $N2$ as the variables. Figures 7(a)–7(d) capture the output corresponding to LPC deficiency as a function of the two input variables. The surface rendered by the Fourier network exhibits more oscillatory slopes relative to the sharply increasing slopes toward the maximum training value of $N2$, which hints at overfitting by the GSLNs and BPFN.

Next, we investigate the noise robustness of the networks by adding a random perturbation to the testing data $\vec{X}$ as follows:

$$X^i_{\text{Noise}} = X^i_{\text{simulated}} + K\sigma^i Z \tag{20}$$

where X^i is the input parameter, σ^i is the parameter sensor's characteristic standard deviation [28], K is the control parameter governing the noise level, and Z a random number drawn from a normal distribution with mean 0 and standard deviation 0.5. Figure 8 shows the results for $0 \leqslant K \leqslant 2$ in steps of 0.1. The accuracy of each GSLN dips sharply with K, and they seem to generalize better at preoptimal epochs, where they achieve a lower-percentage accuracy for noise-free data ($K=0$). This implies overfitting occurring as was apparent in the two-dimensional problem (Fig. 7). The Fourier network outperformed the rest in this respect and clearly presented itself as the most noise-robust five-dimensional model.

Finally, we compare the convergence speeds of the various neural networks. As shown in Fig. 9 and Table 2, the GSLNs possessed a tremendous superiority over the Fourier networks and BPFN in terms of their convergence speeds. They yielded perfect accuracy with the testing data set at around the 20th epoch when the others yielded results of $<50\%$. It is also interesting to observe that the increase in the testing accuracy for Fourier neural networks took place more smoothly than the learning curve for BPFN.

4 Conclusions

We have demonstrated successful applications of both Fourier networks and a class of GSLNs in the two-module fault diagnostics of the F404 engine using a computer simulation program as our data source. With our input choice of five engine parameters—WFE, N2, T5H, PS3, PT56—these neural systems have achieved perfect generalization accuracy when supported by a training-to-testing data size ratio of about 1:5. Similar results hold for conventional BPFN with sigmoidal transfer function. Nonetheless, both novel networks have shown clear superiorities over BPFN in certain aspects, at least for this specific problem. The Fourier network was most resilient against random noise. Essentially, its approximation capability is similar to that of a multidimensional Fourier transform, but with the Fourier coefficients learned via gradient descent. On the other hand, the GSLNs converged to perfect testing accuracy with only ~15 iterations at which the other networks can only yield sub-50% accuracy. What characterizes the GSLN's unique advantage is the usage of Gram-Schmidt Orthonormalization as a self-structuring procedure.

Many avenues exist for future research in these novel neural networks. We have used fixed basis functions throughout training of weight vectors, with their characterizing parameters $\{\vec{\alpha},\vec{a},\vec{b},\beta\}$ in Eq. (15) drawn from probability distributions at each iteration. These parameters determine the shapes of these basis functions, for example, a close-to-zero β for the Gaussian function implies that it has steep slopes. This approach can be extended to one that spends some computational cost to optimize the values of these parameters. In [23], the training algorithm for the GSLN includes, at each epoch, another separate iteration via gradient descent to arrive at the optimal parameters. This compromises the fast convergence speed that comes with fixed-basis functions, but may refine the network's noise robustness. Similarly, for Fourier networks, instead of fixing the first layer weights ($\{W_i\}$ in Fig. 2), we

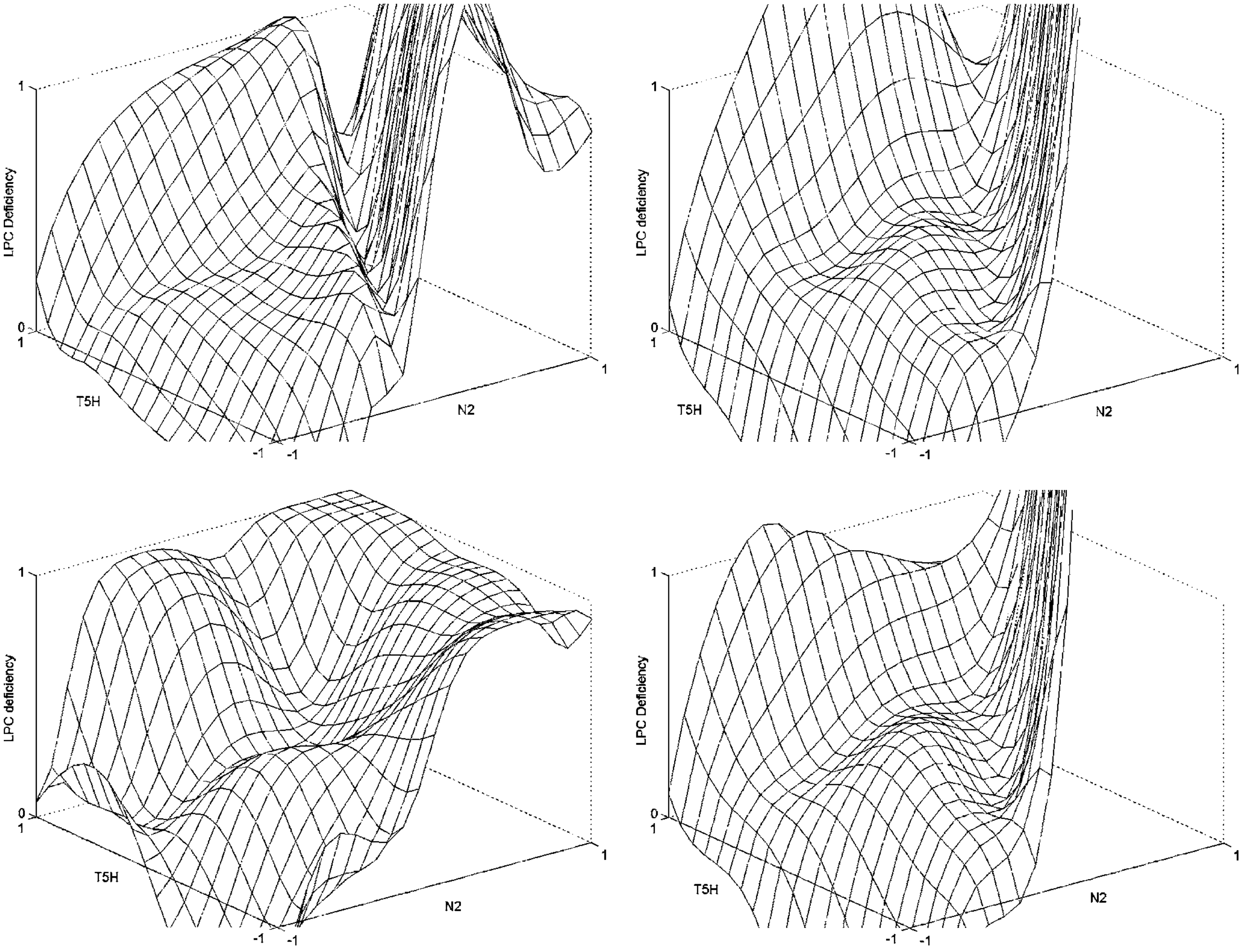


Fig. 7 (*a*)–(*d*) Surface reconstructions by the neural networks

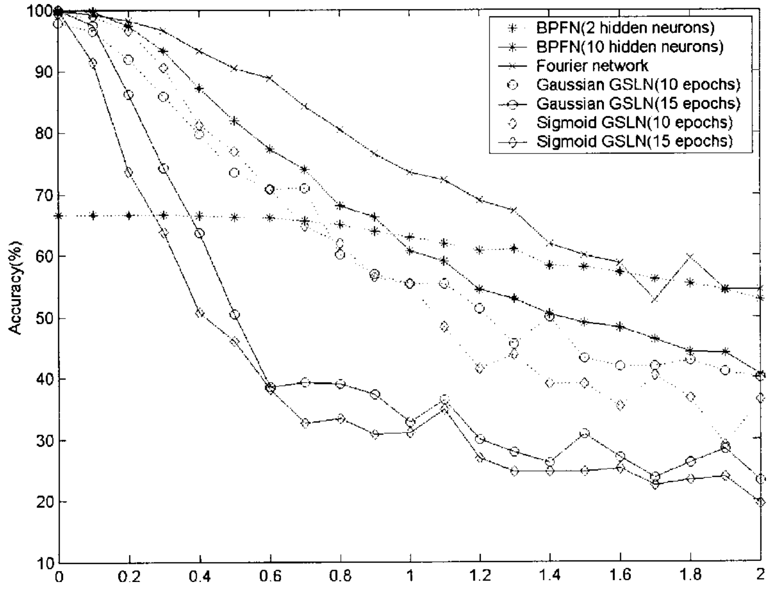


Fig. 8 Noise robustness of various neural networks

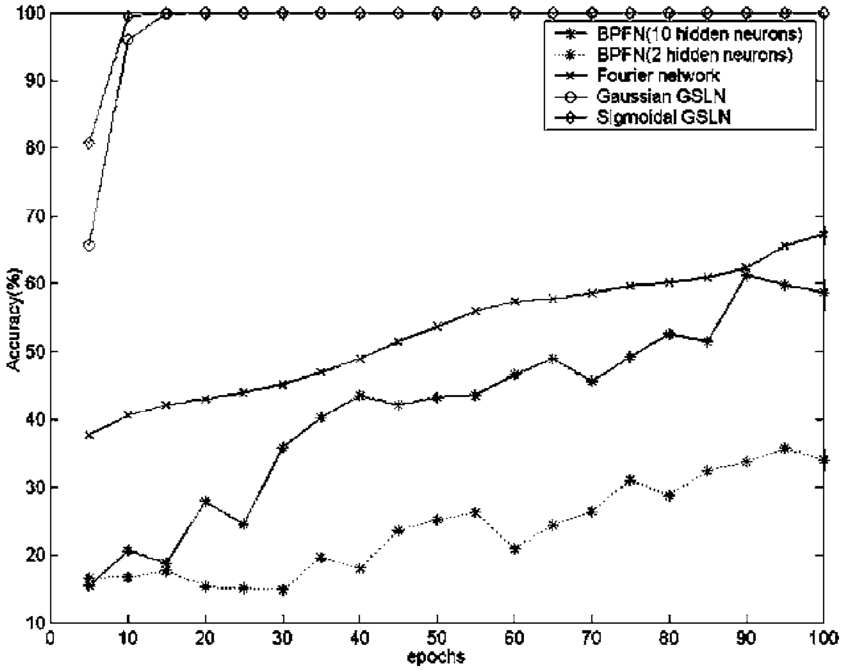


Fig. 9 Networks' accuracies at each epoch

Table 2 Computational time taken for each network to achieve perfect testing accuracy

Neural networks	Fourier network	BPFN (10 hidden neurons)	Gaussian GSLN	Sigmoid GSLN
Computational time (s)	48.5	16	0.75	0.88

can also use gradient-descent methods to calculate them, but this also implies that the basis functions have no strict harmonic relationships as in Fourier transforms.

Currently, data collection is in progress for an eventual real-life application of these neural networks. The applicability of these novel systems lies beyond the steady-state modeling described in this paper. For Fourier networks, the sinusoidal nature of the activation functions renders them useful for time series analysis; as in [29], they were used in the context of recurrent networks to fit temporally varying data. We hope that this work will inspire further use of these neural platforms to a wide range of gas turbine engine diagnostics problems.

Nomenclature

$C(k_1,\ldots,k_D)$ = Fourier coefficients
$F_i(\vec{x})$ = Target output representation of the ith module
K = noise control parameter
U,V,W = weights of neural network
Z = random number from normal distribution with mean zero and standard deviation 0.5
$f(\vec{x})$ = neural network output
$g_i(\vec{x})$ = ith module's deficiency in percentage
$\{\vec{x},Y\}$ = training input-output pairs
ϕ = basis function of GSLN
φ = transfer function of neural network
ψ = orthonormalized basis function of GSLN
μ = fuzzy membership value
σ^i = ith parameter sensor's characteristic standard deviation

References

[1] DePold, H. R., and Gass, F. D., 1999, "The Application of Expert Systems and Neural Networks to Gas Turbine Prognostics and Diagnostics," ASME J. Eng. Gas Turbines Power, **121**, pp. 607–612.
[2] Lu, P. J., Hsu, T. C., Zhang, M. C., and Zhang, J., 2000, "An Evaluation of Engine Faults Diagnostics Using Artificial Neural Networks," ASME Paper No. 2000-GT-0029.
[3] Camporeale, S., Dambrosio, L., Milella, A., Mastrovito, M., and Fortunato, B., 2003, "Fault Diagnosis of Combined Cycle Gas Turbine Components Using Feedforward Neural Networks," ASME Paper No. GT-2003-38742.
[4] Kong, C., Ki, J., Kang, M., and Kho, S., 2004, "Intelligent Performance Diagnostics of a Gas Turbine Engine Using Friendly Interface Neural Networks," Aircraft Eng. Aerospace Tech., **76**, pp. 391–397.
[5] Peng, P., and Yang, M. T., 2000, "Neural Networks Based Diagnostics for Mistuned Bladed Disk," ASME Paper No. 2000-GT-0035.
[6] Sampath, S., Gulati, A., and Singh, R., 2002, "Fault Diagnostics Using Genetic Algorithm for Advanced Cycle Gas Turbine," ASME Paper No. GT-2002-30021.
[7] Sampath, S., Ogaji, S. O. T., Li, Y. G., and Singh, R., 2003, "Fault Diagnosis of a Two Spool Turbo-Fan Engine Using Transient Data: A Genetic Algorithm Approach," ASME Paper No. GT2003-38300.
[8] DePold, H., Chen, D., Ganguli, R., and Volponi, A. J., 2000, "The Use of Kalman Filter and Neural Network Methologies in Gas Turbine Performance Diagnostics: AComparative Study," ASME Paper No. 2000-GT-547.
[9] Hassan, T. A. F., El-Shafei, A., and Zeyada, Y., 2003, "Comparison of Neural Network Architectures for Machinery Fault Diagnosis," ASME Paper No. GT2003-38450.
[10] General Electric, 1996, "F404-GE-100D Cycledeck Program L0046E," GE Aircraft Engine Advance Engineering Program Dept, Lynn, MA.
[11] Tan, H. S., 2004, "Glances at Artificial Intelligence Methods in Aircraft Engine Diagnostics," Republic of Singapore Air Force, Air Logistics Dept., Propulsion Branch Internal Publication.
[12] RSL Electronics Ltd., 2001, *User Manual for Ground Station System for F404-GE-100D Engine Health Monitoring System* .
[13] Kurzke, J., 2003, "Model Based Gas Turbine Parameter Correction," ASME Paper No. GT2003-38234.
[14] Haykin, S., 1999, "*Neural Networks: a Comprehensive Foundation*," Prentice-Hall, Englewood Cliffs, NJ.
[15] Cybenko, G., 1989, "Approximation By Superposition of a Sigmoidal Func-

tion," Math. Control, Signals, Syst., **2**, pp. 303–314.
[16] Wang, Z., Tham, M. T., and Morris, A. J., 1992, "Multilayer Neural Networks: Approximated Canonical Decomposition of Non-Linearity," Int. J. Control, **56**, pp. 655–672.
[17] Silvescu, A., 1999, "Fourier Neural Networks," *Proceedings of the International Joint Conference On Neural Networks*, IEEE, Washington, DC, pp. 488–491.
[18] Kantorovich, L. V., and Akilov, G. P., 1964, *Functional Analysis*, Oxford Pergamon Press, New York.
[19] Walker, J., 1988, *Fourier Analysis*, Oxford University Press, New York.
[20] Zhang, Q., and Benveniste, A., 1992, "Wavelet Networks," IEEE Trans. Neural Netw., **3**, pp. 889–898.
[21] Pittner, S., Kamarthi, S. V., and Gao, G., 1998, "Wavelet Networks for Sensor Signal Classification in Flank Wear Assessment," J. Intell. Manuf., **9**, pp. 315–322.
[22] Strumillo, P., and Kaminski, W., 1997, "Kernal Orthonormalization in Radial Basis Function Neural Networks," IEEE Trans. Neural Netw., **8**, pp. 1177–1183.
[23] Zhang, J., and Morris, A. J., 1998, "A Sequential Learning Approach for Single Hidden Layer Neural Networks," Neural Networks, **11**, pp. 65–80.
[24] Strumillo, P., and Kaminski, W., 2001, *Neural Networks with Orthogonalised Transfer Functions*, ESANN, Belgium.
[25] Duch, W., and Jankowski, N., 1999, "Survey of Neural Transfer Functions," Neural Comput. Surv., **2**, pp. 163–212.
[26] Hagan, M. T., Demuth, H. B., and Beale, M. H., 1996, *Neural Network Design*, PWS Publishing, Boston.
[27] Cover, T. M., 1965, "Geometrical and Statistical Properties of Systems of Linear Inequalities With Applications in Pattern Recognition," IEEE Trans. Electron. Comput., **EC-14**, pp. 326–334.
[28] General Electric Aircraft Engine, 1991, *Technical Manual Intermediate Maintenance Turbofan Engine F404-GE-100D Model*, SEI-701.
[29] Koplon, R., and Sontag, E. D., 1997, "Using Fourier Neural Recurrent Networks to Fit Sequential Input/Output Data," Neurocomputing, **15**, pp. 225–248.

A Reassessment of the Alternative Regeneration Cycle

Paul A. Dellenback
Associate Professor
Department of Mechanical Engineering,
University of Wyoming,
Laramie, WY 82072-3295
e-mail: pad@uwyo.edu

Two prior papers and several patents have considered improvements to a gas turbine engine's cycle efficiency by using two turbines in series with an intermediate heat exchanger that preheats combustion air. This approach allows heating the combustion air to temperatures higher than those that can be achieved with "conventional regeneration" in which the combustion products are fully expanded across a turbine before any heat recovery. Since heat addition in the combustor of the "alternative regeneration" cycle occurs at a higher average temperature, then under certain conditions the cycle efficiency can be higher than that available from a cycle using conventional regeneration. This paper reconsiders the usefulness of the alternative regeneration cycle with more detailed modeling than has been presented previously. The revised modeling shows that the alternative regeneration cycle can produce efficiencies higher than conventional regeneration, but only for a more limited set of conditions than previously reported. For high-technology engines operating at high temperatures, the alternative regeneration cycle efficiencies can be three to four percentage points better than comparable conventional regeneration cycles. For lower-technology engines, which are more typical of those currently installed, improvements in efficiency only occur at lower values of heat exchanger effectiveness, which limits the usefulness of the alternative regeneration cycle. Also considered is an extension to the cycle that employs a second heat exchanger downstream of the second turbine for the purpose of further preheating the combustion air. In its optimum configuration, this "staged heat recovery" can produce additional small improvements of between 0.3 and 2.3 percentage points in cycle efficiency, depending on the particular cycle parameters assumed. [DOI: 10.1115/1.2179079]

Introduction

The alternative regeneration scheme shown in Fig. 1(*c*) was discussed in prior papers by Dellenback [1] and Cardu and Baica [2]. The cycle utilizes a high-pressure turbine (HPT), a power turbine (PT), and a heat exchanger located between the two turbines. Heat is extracted from the hot gas stream leaving the first turbine and used to preheat air passing to the combustor. Because the hot gases from the HPT are not fully expanded before entering the heat exchanger, their temperature is higher than would be the case in a conventional regeneration scheme (see Fig. 1(*b*)) where heat is recovered only after the maximum amount of work has been extracted. There are two competing consequences associated with using the underexpanded gases for preheating air passing to the combustor: (i) heat addition at the combustor occurs at a higher average temperature, which tends to increase cycle efficiency, and (ii) the specific power output of the cycle is reduced because the enthalpy of gases passing to the PT is lower. Whether the overall cycle efficiency is increased or decreased by the alternative heat exchanger location depends on which of the two consequences dominates and this, in turn, is a function of cycle operating parameters. Dellenback [1] shows how simple thermodynamic models predict significant efficiency gains for the alternative regeneration cycle over a wide variety of operating conditions.

In addition to the reduced specific power output, a second drawback of the alternative regeneration cycle is that the heat exchanger operates at higher temperatures and pressures than is the case for a conventional regeneration cycle. The pressure requirements are higher because the optimum pressure ratio (PR) for the alternative regeneration cycle tends to be higher than for conventional regeneration. However, both the temperature and pressure requirements are within the capabilities of modern heat exchangers, as discussed by Wright and Stringer [3].

Figure 2 contrasts a simple cycle, a conventional regenerative cycle, and the alternative regeneration cycle, with all operating between the same temperature and pressure limits. The heat exchanger is assumed to be a recuperator that has no moving parts and no gas leakage. Since the heat input is equivalent to the area under the curve on a *T*-*s* diagram, the dramatically different levels of external heat input at the combustor are readily apparent in Fig. 2 when comparing paths 2 to 4 for the simple cycle, path 3_{conv} to 4 for conventional regeneration, and path 3_{alt} to 4 for alternative regeneration. The power output from the cycles is proportional to the area enclosed within the curves, so that the smaller values of specific power produced by the alternative regeneration cycle are also apparent. The heat rejected from the cycles is proportional to the area under the curves; from states 6 to 1 for the simple cycle, from 7_{conv} to 1 for the conventional regeneration cycle, and from 7_{alt} to 1 for the alternative regeneration cycle. Of the three cycles, the alternative regeneration cycle requires the smallest heat input, produces the lowest specific power, and rejects the smallest amount of heat to the environment.

When the alternative regeneration cycle is operated as a two-spool engine with gas generator and PT utilizing separate shafts, the cycle efficiency will be less than optimum. Hence, Fig. 1(*c*) shows a single-shaft configuration that allows optimization of the efficiency by appropriate choice of the pressure between the turbines. The optimum configuration of the alternative regeneration cycle occurs when the PT is supplying some fraction of the net power output. Note that in the limit as the PT output goes to zero, the alternative regeneration cycle is exactly the same as the conventional regeneration cycle.

The paper by Dellenback [1] uses an elementary approach for developing comparative thermodynamic cycle models, and within the limits of those models, the alternative regeneration scheme exhibits attractive characteristics. The goal of the present work was to upgrade the sophistication of the original thermodynamic

---

Contributed by the Electric and Power Committee of the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received March 9, 2004; final manuscript received August 19, 2005. Review conducted by S. van der Linden.

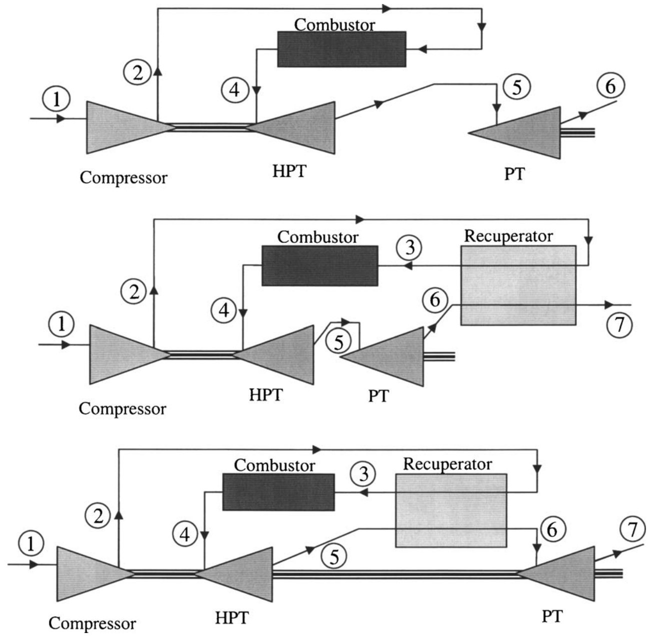


Fig. 1 Configuration of simple, conventional regeneration, and alternative regeneration cycles: (*a*) simple cycle in gas generator configuration, (*b*) conventional regeneration in gas generator configuration, (*c*) alternative regeneration cycle

models, and this, generally, had a negative impact on the attractiveness of the alternative regeneration cycle. The alternative regeneration cycle models are particularly sensitive to the way properties are evaluated, so that while the results presented in Dellenback [1] are accurate, they are also overly optimistic with respect to the advantages of alternative regeneration when contrasted with conventional regeneration. Finally, the benefits of adding a second heat exchanger downstream of the PT were examined. Although this would not seem to be feasible for the particular temperature and pressure limits assumed in Fig. 2 (because $T_{7\mathrm{alt}} < T_2$), there are other pressure and temperature conditions for which this second heat exchanger is feasible, with a resulting benefit for cycle efficiency.

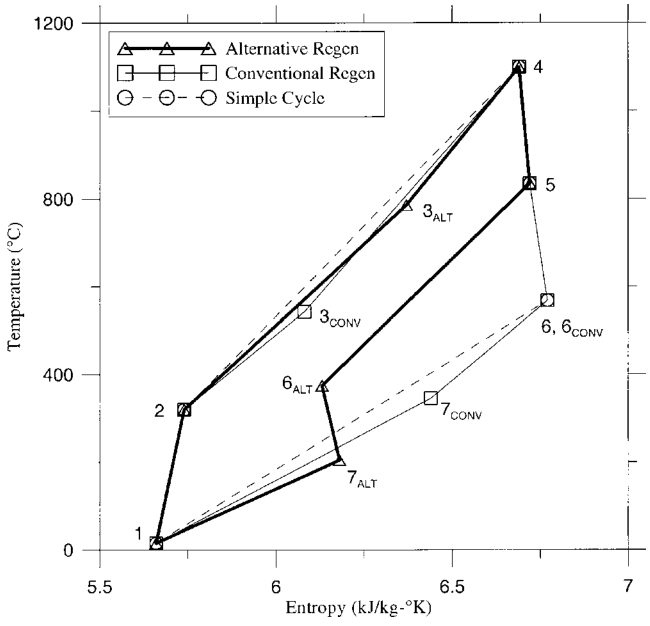


Fig. 2 Comparison of cycles for PR=10, TIT=1100°C, effectiveness=0.9, $\Delta P_{\mathrm{rec}}=0$

Thermodynamic Models

To improve the sophistication of the thermodynamic modeling, two aspects of the original models [1] were modified and applied to all calculations shown in the present paper:

(a) Polytropic compressor (η_{pc}) and turbine (η_{pt}) efficiencies were incorporated in place of the isentropic efficiencies used in the original paper. This change had a negligible impact on the results shown in Dellenback [1], but it is, nevertheless, a more correct approach for comparing cycles having different pressure ratios [4,5].

(b) Pressure drops through the combustor and heat exchanger have been modeled as a percent of inlet pressure to these devices, as opposed to an absolute value of 13.8 kPa (2 psi) that was used in Dellenback [1]. This revised approach to modeling the pressure drops tends to penalize the alternative regeneration cycle significantly more than the conventional regeneration cycle because the hot-side pressure drop penalty in the recuperator is more severe than for conventional regeneration, resulting in less work extraction from the PT.

For the simple and conventional regenerative cycles, so long as polytropic efficiencies are used, it is not important from a thermodynamic modeling perspective whether the engine is configured as a gas generator with PT or as a single-shaft machine having one compressor and only one turbine. This is not true of the alternative regeneration cycle since heat is extracted between the two turbines.

For the states as labeled in Fig. 1, when an air-standard cycle is assumed, the following thermodynamic models result:.

Simple Cycle:

$$\eta_{\mathrm{cycle,simple}} = \frac{\text{Net work output}}{\text{External heat input}} = \frac{\dot{m}(h_5 - h_6)}{\dot{m}(h_4 - h_2)} = \frac{(h_5 - h_6)}{(h_4 - h_2)} \quad (1)$$

Conventional Regeneration:

$$\eta_{\mathrm{cycle,conv\ regen}} = \frac{(h_5 - h_6)}{(h_4 - h_3)} \quad (2)$$

Alternative Regeneration:

$$\eta_{\mathrm{cycle,alt\ regen}} = \frac{(h_6 - h_7) + (h_4 - h_5) - (h_2 - h_1)}{(h_4 - h_3)} \quad (3)$$

The form of Eq. (3) allows for modeling the cycle as either a gas generator with PT (for which $(h_4 - h_5) = (h_2 - h_1)$) or as a single-shaft machine with the two turbines and compressor all connected on a common shaft. There has been no allowance for turbine blade cooling in these models. To evaluate temperatures after compression, the polytropic efficiency is used to give

$$\frac{T_2}{T_1} = \left(\frac{P_2}{P_1}\right)^{(\gamma-1)/\gamma\eta_{\mathrm{pc}}} \quad (4)$$

and for a typical expansion (e.g., for the alternative regeneration cycle), the downstream temperature is found from

$$\frac{T_7}{T_6} = \left(\frac{P_7}{P_6}\right)^{(\gamma-1)\eta_{\mathrm{pt}}/\gamma} \quad (5)$$

The definition of heat exchanger effectiveness is used to find the temperatures at state 3, and energy balances on the heat exchanger give temperatures at 7 and 6 in the conventional and alternative regeneration cycles, respectively. For each cycle, at each operating condition, the compressor pressure ratio (PR) was optimized to find the maximum cycle efficiency (an exception was the Fig. 2 data where PR=10 was used to demonstrate the differ-

Table 1 Model parameters

Compressor inlet temperature	15 °C
Compressor inlet pressure	1 atm
Combustor pressure drop	3% of inlet pressure
Heat exchanger pressure drops	2% of inlet pressure for each gas path
Polytropic compressor efficiency	90%
Polytropic turbine efficiency	87%

ence between the three cycles). In addition, for the single-shaft configuration of the alternative regeneration cycle, the pressure at state 5 was simultaneously optimized to give the highest cycle efficiency. Unless stated otherwise, all models used the parameters shown in Table 1.

In order to solve Eqs. (1)–(5), a method of relating temperatures to enthalpies is required. The alternative regeneration cycle proved particularly sensitive to the property evaluation scheme that was used. Temperature-dependent specific heats were used, which makes the calculation iterative since the temperatures required for property evaluation are not known a priori. ENGINEERING EQUATION SOLVER (EES) [6] was used as the modeling platform to facilitate both the property evaluation and the iterative solution.

Both specific heats (at constant volume and constant pressure) were evaluated at each state and then an average value of the specific heat ratio was evaluated at the mean temperature associated with each compression and expansion for use in Eqs. (4) and (5), respectively. The temperature-dependent specific heat data embedded in EES are based on the gas tables by Keenan et al. [7]. The working fluid was taken to be pure air, behaving as an ideal gas.

A second approach for evaluation of enthalpy was also considered, where product gases were modeled as mixtures of oxygen, nitrogen, carbon dioxide, and water vapor. The mixture composition was a function of the fuel/air ratio required to give the desired turbine inlet temperature (TIT). This approach was not an air-standard cycle analysis because the fuel mass was included in the models and Eqs. (1)–(3) modified accordingly. This scheme was examined for a wide variety of operating conditions, but produced results essentially identical to the cases where product gases were modeled as pure air. The pure air results are presented here since the calculations were simpler to implement and devoid of any convergence difficulties that sometimes arose in the multiple-parameter optimization for overall PR and PT inlet pressure.

Results and Discussion

The curves in Fig. 3 represent the range of turbine inlet temperature from 900 °C to 1500 °C, with PR optimized at each point to give the highest possible cycle efficiency. Although TITs of 1500 °C are rather high for current technology, the goal of this investigation was to examine both the present and future feasibility of the cycle. Results for two configurations of the alternative regeneration cycle are shown, one being the single-shaft configuration with optimized pressure at state 5, and the other the gas generator configuration where the HPT work is set equal to the compressor work requirement. When compared to the results in Dellenback [1], where a simpler modeling approach was used, we see that those prior models significantly underpredict all three cycle efficiencies compared to the more accurate model results shown here. For the simple cycle at TIT of 1300 °C and higher, the optimum PR is excessive by current standards; thus, the PR was limited to 40 and this explains the declining slope of the curve at higher TIT. The Fig. 3 results show that alternative regeneration will be superior to conventional regeneration under these conditions, and the advantage increases with increasing TIT. This is due to the fact that the average temperature at which the PT work is being extracted is lower than with conventional regeneration. This lower temperature leads to a higher value of the specific heat for product gases (2% higher across the alternative regeneration PT) and thus, an increase in cycle performance. However, the benefits of the alternative regeneration are only realized for lower recuperator effectivenesses (>0.9)—at the higher effectivenesses typically realized in modern recuperators and regenerators, the conventional regeneration cycle will give superior performance.

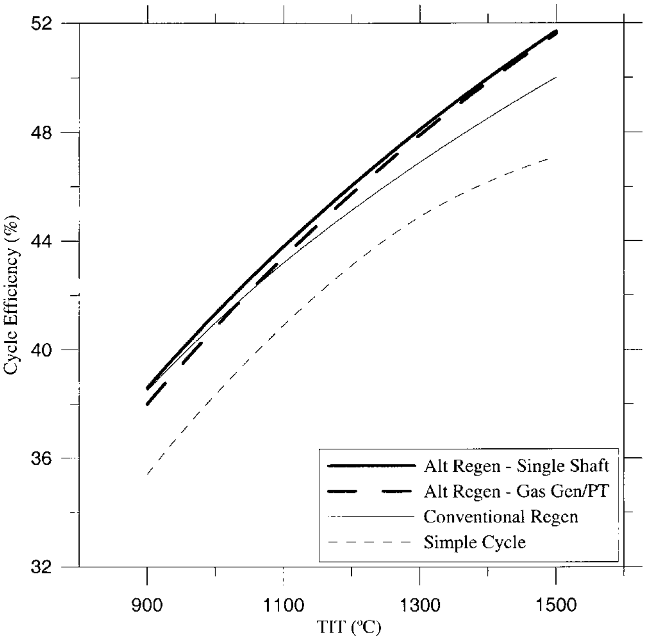


Fig. 3 Comparative cycle performance, effectiveness=0.7; PR optimized for each point

For the single-shaft alternative regeneration cycle data of Fig. 3, the PT contribution to the net shaft power varies from 33% at a TIT of 900 °C to 81% at a TIT of 1500 °C. In later discussion, it will be seen that for some configurations, the PT is producing a much smaller fraction of the total shaft work.

The cycle PR is obviously important in engine design since higher PRs require larger and more expensive compressors. Figure 4 shows the optimum PRs corresponding to the cases in Fig. 3. Again, the simple-cycle PR was limited to 40 or less. In addition to the increased cycle efficiency afforded by conventional regeneration, Fig. 4 shows that the optimum PR is much smaller than for the simple cycle. The improved efficiencies available from the alternative regeneration cycles are only available by using PRs higher than those required for conventional regeneration, but these higher optimum PRs are feasible for current engines.

Figure 5 shows the PT exit temperature, giving an indication of useful heat that remains in the exhaust gases. The alternative regeneration cycles provide more heating to the combustion air than conventional regeneration, thus lowering the temperature of gases through their PTs and also lowering the PT exit temperature.

Feasibility Limits for the Alternative Regeneration Cycle. Figure 3 gave results for a recuperator effectiveness of 70% to show that the alternative regeneration cycle can, theoretically, offer more attractive performance than a simple or conventional regeneration cycle. Since the use of heat-recovery heat exchangers with gas turbines has been limited both by the bulk or size of the exchanger as well as the increased cost, the lower effectivenesses might be preferred in some space or cost-sensitive applications because even a low-effectiveness heat exchanger will give cycle efficiencies significantly higher than a simple cycle, as shown by Fig. 6. However, current state-of-the-art heat exchanger designs, whether of the recuperator or regenerator type, typically have effectivenesses in excess of 90%. Recall that the single-shaft alter-

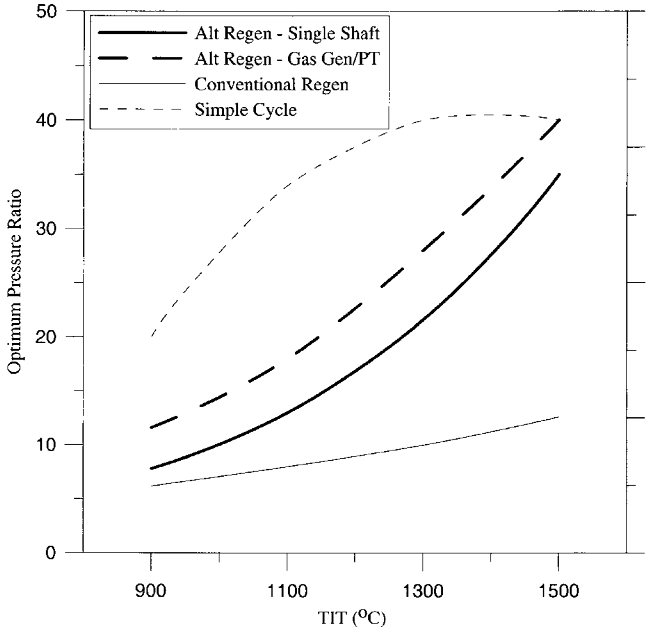


Fig. 4 Optimum pressure ratios for results shown in Fig. 3 (limited to a maximum value of 40)

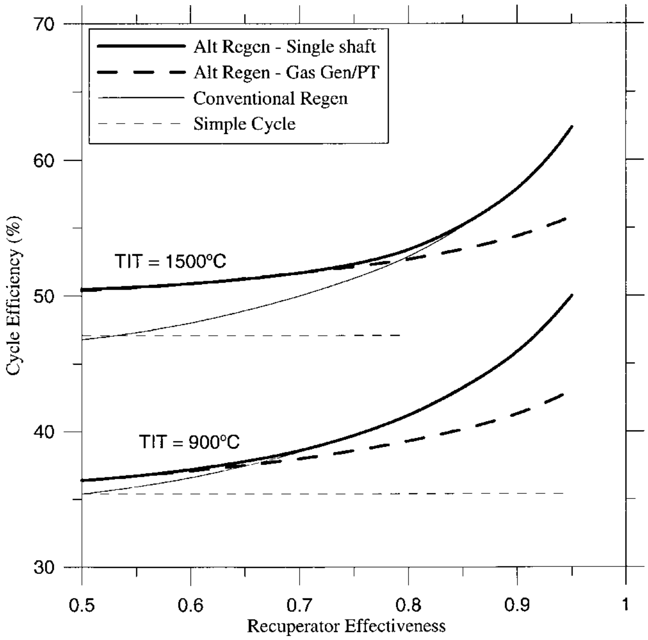


Fig. 6 Effect of recuperator effectiveness at two different TITs, PR optimized for each operating point

native regeneration cycle results were optimized as a function of two parameters—the overall PR and the pressure between the two turbines. At the higher effectivenesses, the best-case performance of the alternative regeneration cycle corresponds to the limiting case where all of the gas expansion occurs across the HPT such that the alternative regeneration cycle is identical to conventional regeneration. However, when the effectiveness is low, the higher combustor inlet temperatures available from the alternative regeneration cycle act to offset the lower PT performance, giving an advantage over conventional regeneration. Figure 6 shows how the cycle efficiency varies as a function of effectiveness, with the upper four curves corresponding to TIT=1500°C and the lower four curves corresponding to TIT=900°C. At 1500°C, alternative regeneration has no benefits over conventional regeneration for effectivenesses greater than about 0.83, and for the lower TIT of 900°C, the feasibility limit for alternative regeneration occurs at an effectiveness of ~0.68. If alternative regeneration is to be used in a gas generator/PT configuration, then Fig. 6 shows that this configuration is only better than conventional regeneration for an even more limited range of effectivenesses.

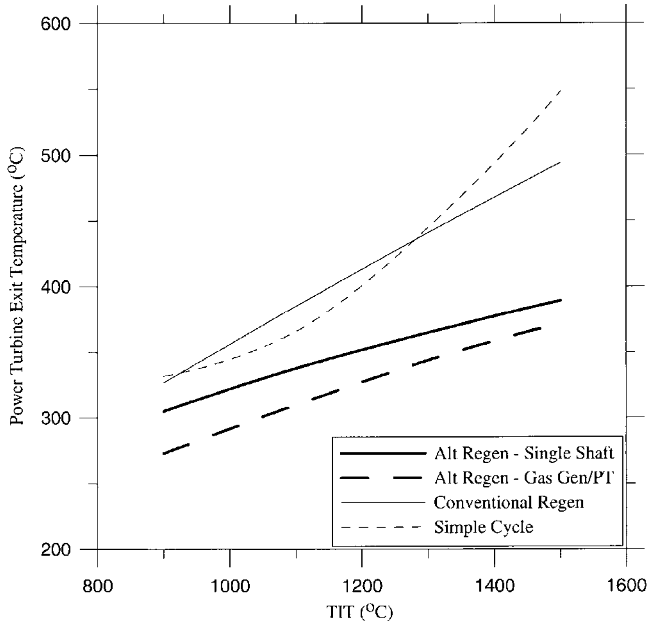


Fig. 5 Power turbine exit temperatures for results shown in Fig. 3

In heat exchanger design, the cross-sectional area of flow channels can be increased with the goal of lowering the pressure drop, with the consequence of lower effectiveness if the overall size and cost of the heat exchanger is fixed. Lowering the heat exchanger pressure drop can also cause problems in obtaining uniform flow distribution through the exchanger [8]. Nevertheless, as pressure drop through the heat exchanger is reduced, there would seem to be greater benefit for the alternative regeneration cycle than for conventional regeneration since the hot side gases are at higher pressure in the alternative regeneration cycle and the pressure loss is evaluated as 2% of the incoming pressure. Thus, a heat exchanger with low pressure drop and low effectiveness might help to improve the range of useful applications for the alternative regeneration cycle. However, the overall cycle performance is much more sensitive to the effectiveness than the heat exchanger pressure drop. Thus, calculations revealed no substantial improvement in the range of feasibility for the alternative regeneration cycle by using a heat exchanger with low effectiveness and very low pressure drop.

Staged Heat Recovery. For many of the alternative regeneration operating points already presented, the gas temperature out of the PT was higher than the compressor outlet temperature (i.e., $T_7 > T_2$). This suggests there could be a benefit by adding a second heat exchanger, in a "staged heat recovery" (SHR) scheme. This second exchanger would use gases discharged by the power turbine to preheat air between the compressor and primary recuperator as shown in Fig. 7. Since such a heat exchanger would introduce additional pressure drops into both hot- and cold-side gas flow paths, the feasibility of the SHR cycle depends on whether enough heat can be recovered to offset the increased pressure drop penalty. The SHR scenario was modeled by using the definition of heat exchanger effectiveness to determine the outgoing air temperature (T_9), and an energy balance on the heat ex-

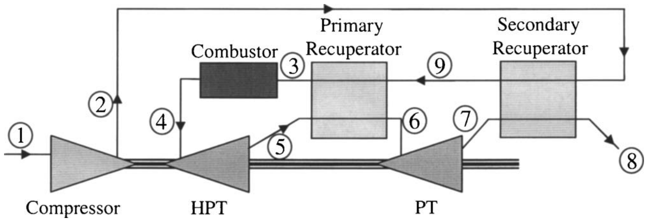


Fig. 7 Configuration of the staged heat recovery (SHR) cycle

changer to determine the final exhaust temperature of the combustion gases (T_8). The pressure drops for the second heat exchanger were taken to be equal to the values assumed for the primary recuperator.

The addition of a second heat exchanger in the SHR configuration proved to be feasible only at high TIT. For a TIT of 1500°C, Fig. 8 shows that the simple cycle efficiency peaks at PRs that are higher than the maximum feasible value for state-of-the-art engines, which is ~40. The conventional regeneration cycle offers increased efficiency at a modest PR of ~13. Alternative regeneration can further improve the cycle efficiency by ~1.7 percentage points, but at the expense of a larger compressor providing a PR of ~30. Finally, the SHR cycle can further improve the cycle efficiency and can do so at a much more manageable PR of 9. The reason the SHR cycle reduces the optimum PR is because the PT is only producing ~12% of the total net power output at the peak efficiency operating point; hence, the cycle is operating very similarly to a conventional regeneration cycle. When a second heat exchanger is added to a conventional regeneration cycle, there is also an improvement in cycle efficiency because the overall effect is to recover more useful heat from the exhaust gas stream, just as might be done with a single heat exchanger having very high effectiveness. Although the PT is producing a small fraction of the total work in the optimum SHR cycle, there is a slight improvement of ~0.2 percentage points in cycle efficiency over that yielded by a conventional regeneration cycle using two heat exchangers in series.

The Fig. 8 results were generated assuming the effectiveness of both heat exchangers was the same. There is the potential to fine tune the SHR cycle by manipulating the effectivenesses, or the quantity of heat transferred, by each exchanger. This was examined by varying each effectiveness independent of the other. Although the cycle efficiencies increased as either effectiveness was increased, the SHR models proved to be insensitive to the particular order in which heat was removed. For example, if the effectiveness was 0.7 for the primary recuperator and 0.9 for the secondary recuperator, the results were essentially identical when the two effectiveness values were interchanged. This is apparently due to the aforementioned observation that the PT is performing very little work in the optimized SHR cycle so that subtle changes in the PT inlet temperature do not have a significant influence on the overall cycle efficiency.

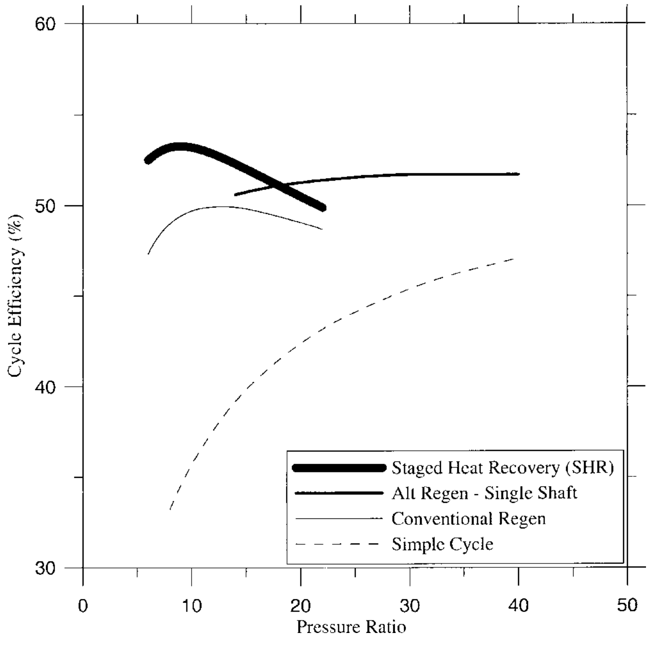


Fig. 8 Optimum PR is a strong function of cycle configuration, effectiveness=0.7, TIT=1500°C

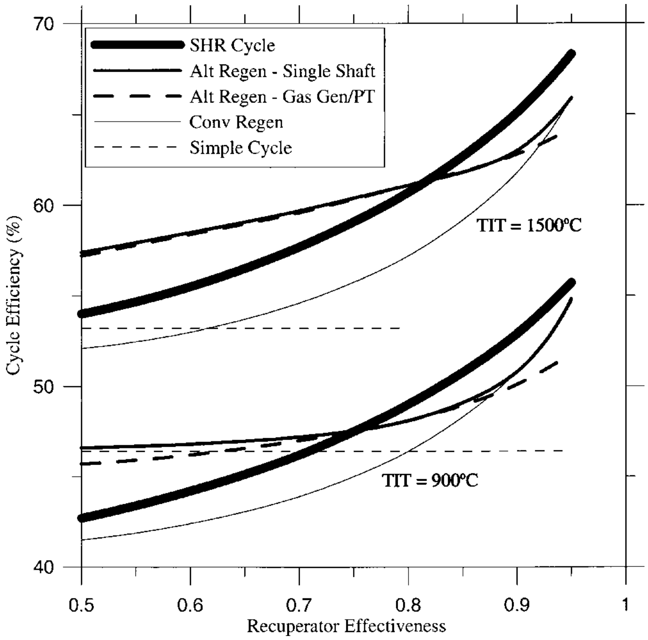


Fig. 9 Effectiveness range for which alternative regeneration is feasible is expanded when higher technology components are available; η_{pc}=0.95, η_{pt}=0.92. PR optimized for each point, but limited to 40 or less.

The SHR cycle proved to be feasible at high values of heat exchanger effectiveness, at least for high TIT. At a TIT of 1500°C, the SHR cycle consistently produced a cycle efficiency ~0.2 percentage points better than a conventional regeneration cycle using two heat exchangers in series for values of effectiveness up to 0.95. The SHR cycle was the only version of alternative regeneration that produced some benefit in cycle efficiency (relative to conventional regeneration) for high values of effectiveness.

The performance of the cycles that have been modeled here are quite sensitive to the compressor and turbine efficiencies. As technology improves and component efficiencies increase, the alternative regeneration cycles will be more attractive relative to conventional regeneration and simple cycles. To illustrate this point, the Fig. 6 calculations were reconsidered with compressor polytropic efficiency increased from 0.90 to 0.95, and with turbine polytropic efficiency increased from 0.87 to 0.92. The results, shown in Fig. 9, indicate that the feasibility of the alternative regeneration cycles is extended to higher effectivenesses as the component efficiencies increase. For a TIT of 1500°C and effectiveness of 0.9, single-shaft and gas-generator versions of alternative regeneration have cycle efficiencies that are ~1 percentage points higher than conventional regeneration. The improvement is substantially larger at lower effectivenesses. The SHR cycle appears to be particularly attractive at high effectiveness. For example, at an effectiveness of 0.9 and TIT of 1500°C, the SHR cycle efficiency is ~3.3 percentage points better than that of conventional regeneration. However, at this condition, the PT in the SHR cycle is once again providing very little of the net power output $(\sim 7\%)$ so that the optimum SHR cycle is again very much like a conventional regeneration cycle having two heat exchangers in series. In fact, the SHR cycle efficiency is only ~ 0.3 percentage points better than a conventional regeneration cycle using two heat exchangers when the effectiveness is 0.90 and $\text{TIT}=1500°\text{C}$. Hence, the primary conclusion from Fig. 9 is that the alternative regeneration cycles using one heat exchanger will be more feasible for future high-technology engines that have high component efficiencies, and the SHR cycle will be of little interest.

Conclusions

The alternative regeneration cycle has been modeled with more detailed models than have been presented previously. The results are significantly different from those given in an earlier paper, in part, due to the way thermodynamic properties were evaluated. The alternative regeneration cycle exhibits sensitivity to the property evaluation scheme because the cycle extracts work over a larger range of temperatures than either a simple or conventional regeneration cycle operating between the same temperature and pressure limits, and therefore property variation is more extreme. The alternative regeneration cycle can exhibit performance superior to conventional regeneration, but this usually only occurs for lower values of heat exchanger effectiveness so that the usefulness of the cycle is limited. As component efficiencies increase, alternative regeneration is attractive at higher effectivenesses. The optimum pressure ratios for alternative regeneration cycles tend to be higher than for conventional regeneration, requiring larger and more expensive compressors to achieve a particular level of power output. When a second heat exchanger is added to the alternative regeneration cycle, there is a small benefit for cycle efficiency achievable at higher effectivenesses. When using a second heat exchanger, the optimum PRs are much lower and virtually the same as those required for conventional regeneration.

Nomenclature

h = enthalpy
HPT = high-pressure turbine
$\dot{m}$ = mass flow rate
P = pressure
PR = compressor pressure ratio
PT = power turbine
SHR = staged heat recovery
T = temperature
TIT = turbine inlet temperature

Greek

η = cycle efficiency
γ = specific heat ratio

Subscripts

1,2,3,... = thermodynamic states identified in Figs. 1 and 7
pc = polytropic compressor efficiency
pt = polytropic turbine efficiency
rec = recuperator

References

[1] Dellenback, P. A., 2002, “Improved Gas Turbine Efficiency Through Alternative Regenerator Configuration,” ASME J. Eng. Gas Turbines Power, **124**(3), pp. 441–446.
[2] Cardu, M., and Baica, M., 2002, “Gas Turbine Installations With Divided Expansion,” Energy Convers. Manage., **43**, pp. 1747–1756.
[3] Wright, I. G., and Stringer, J., 1997, “Materials Issues for High-Temperature Components in Indirectly Fired Cycles,” ASME Paper No. 97-GT-300.
[4] Wilson, D. G., 1984, *The Design of High-Efficiency Turbomachinery and Gas Turbines*, MIT Press, Cambridge, MA.
[5] Crumpsty, N., 1997, *Jet Propulsion*, Cambridge University Press, Cambridge, England.
[6] Engineering Equation Solver©, 2002, S. A. Klein, F-Chart Software, Madison, WI.
[7] Keenan, J. H., Chao, J., and Kaye, J., 1983, *Gas Tables: Thermodynamic Properties of Air Products of Combustion and Component Gases*, Wiley, New York.
[8] Beck, D. S., and Wilson, D. G., 1996, *Gas Turbine Regenerators*, Chapman and Hall, New York.

An Adaptation Approach for Gas Turbine Design-Point Performance Simulation

Y. G. Li
P. Pilidis
School of Engineering,
Cranfield University,
Bedford MK43 0AL, UK

M. A. Newby
Manx Electricity Authority,
Isle of Man, UK

Accurate simulation and understanding of gas turbine performance is very useful for gas turbine users. Such a simulation and performance analysis must start from a design point. When some of the engine component parameters for an existing engine are not available, they must be estimated in order that the performance analysis can be carried out. However, the initially simulated design-point performance of the engine using estimated engine component parameters may give a result that is different from the actual measured performance. This difference may be reduced with better estimation of these unknown component parameters. However, this can become a difficult task for performance engineers, let alone those without enough engine performance knowledge and experience, when the number of design-point component parameters and the number of measurable/target performance parameters become large. In this paper, a gas turbine design-point performance adaptation approach has been developed to best estimate the unknown design-point component parameters and match the available design-point engine measurable/target performance. In the approach, the initially unknown component parameters may be compressor pressure ratios and efficiencies, turbine entry temperature, turbine efficiencies, air mass flow rate, cooling flows, bypass ratio, etc. The engine target (measurable) performance parameters may be thrust and specific fuel consumption for aero engines, shaft power and thermal efficiency for industrial engines, gas path pressures and temperatures, etc. To select, initially, the design point component parameters, a bar chart has been used to analyze the sensitivity of the engine target performance parameters to the design-point component parameters. The developed adaptation approach has been applied to a design-point performance matching problem of an industrial gas turbine engine GE LM2500+ *operating in Manx Electricity Authority (MEA), UK. The application shows that the adaptation approach is very effective and fast to produce a set of design-point component parameters of a model engine that matches the actual engine performance very well. Theoretically, the developed techniques can be applied to other gas turbine engines.* [DOI: 10.1115/1.2136369]

Introduction

Thermodynamic models and computer software for gas turbine performance analysis have been available for a long time. When all engine internal design-point component parameters such as airflow rate, compressor pressure ratios and isentropic efficiencies, turbine entry temperature, turbine efficiencies, cooling flows, etc., are available, engine overall design-point performance (such as thrust or shaft power, SFC or thermal efficiency, gas path pressures and temperatures, etc.) can be predicted very successfully. On the contrary, if some of the component parameters are not available, they must be estimated in order that engine performance can be predicted. However, the predicted engine performance may be slightly different from real performance. This is especially true for gas turbine users who want to know their engine performance but cannot access the engine manufacturer's data. On the other hand, the performance of individual engines in the same fleet may be slightly different from engine to engine due to manufacturing and assembly tolerance. For gas turbine diagnostics engineers, the performance difference in fleet engines may be significantly important when analyzing the performance degradation for an individual engine. Therefore, accurate performance simulation for individual engines is highly demanded.

Gas turbine off-design performance adaptation methods for gas path diagnostic purpose have been explored by many researchers. Stamatis et al. [1] introduced modification factors (MF) and used a nonlinear generalized minimum residual method to optimize the MFs. Lambiris et al. [2] introduced a weighted error function and used a polytope algorithm to optimize the MFs. Stamatis et al. [3] introduced a sensitivity measure and a fast-selection procedure based on the method of singular value decomposition for the optimization of MFs. Kong et al. used a scaling method [4] and a compressor map generation method with genetic algorithms [5] for engine performance adaptation. All the above-mentioned performance adaptation methods at off-design or degraded conditions assume that the simulated performance at design point match real performance very well. Such a design-point adaptation or matching is normally done by proper selection of a set of design-point component parameters with trial and improvement. In practice, the design-point performance for engines with complex configurations may be difficult to match due to lack of enough component data. In other words, they may have too many design-point component parameters to guess in order to match the engine design-point performance. Recently, Roth et al. introduced a Bayesian updating and inference concept [6] and an optimization concept [7] for engine-cycle model matching and the idea is tested in a simple surrogate matching problem. More work is to be done to apply the idea to gas turbine engines.

A gas path analysis (GPA) for gas turbine diagnostics used by Escher and Singh [8] is based on the iterative use of the inverse of fault coefficient matrix (FCM) to estimate component degradation corresponding to the deviation of measurable performance param-

---

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the Journal of Engineering for Gas Turbines and Power. Manuscript received August 25, 2005; final manuscript received September 28, 2005. IGTI Review Chair: K. C. Hall. Paper presented at the ASME Turbo Expo 2005: Land, Sea, and Air, Reno, NV, June 6–9, 2005, Paper No. GT2005-68140.

eters. Although the method was initially developed for the analysis of gas turbine degradation, the mathematical model can be extended to gas turbine design-point performance adaptation.

In this paper, a novel linear and a nonlinear design-point performance adaptation approach is introduced, where a linearized engine thermodynamic relationship and a matrix inverse method are used. The approach is applied to the design-point performance matching of an industrial gas turbine engine GE LM2500+ running at Manx Electricity Authority (MEA), UK. The adaptation results produced with different model engine configurations, different adaptation setting, and different adaptation methods are provided and analyzed. The advantages and disadvantages of the approach are discussed thereafter.

Mathematical Model for Adaptation

Design-point performance adaptation is an inverse mathematical problem. In other words, the information of the dependent variables in a system is used to estimate the quantity of the independent variables in the system. The matrix inverse mathematical model used in the GPA approach used by Escher and Singh [8] in their gas turbine diagnostics is an effective solver to model-based inverse problems and can be applied to engine design-point performance adaptation. The idea and the mathematical representation of the approach are described as follows.

In design-point performance adaptation process, two types of adaptation parameters are defined and used:

- To-be-adapted component parameters—they are also called independent adaptation parameters or gas turbine design-point component parameters, such as airflow rate, compressor pressure ratios and isentropic efficiencies, turbine entry temperature, turbine efficiencies, cooling flows, etc.
- Target performance parameters—they are also called dependent adaptation parameters, measurable or measured performance parameters, such as thrust or shaft power, SFC or thermal efficiency, gas path pressures and temperatures, etc.

The thermodynamic relationship between gas turbine to-be-adapted component parameters and target performance parameters of a gas turbine engine can be expressed with

$$\vec{z} = h(\vec{x}) \tag{1}$$

where $\vec{z} \in R^M$ is the target performance parameter vector and M the number of the target performance parameters, $\vec{x} \in R^N$ is the to-be-adapted component parameter vector and N the number of to-be-adapted component parameters, and $h(\ )$ is a vector-valued function. At a given initial design point (baseline), which is denoted by subscript 0, Eq. (1) can be expanded in a Taylor series

$$\vec{z} = \vec{z}_0 + \left.\frac{\partial h(\vec{x})}{\partial \vec{x}}\right|_0 (\vec{x} - \vec{x}_0) + \text{HOT} \tag{2}$$

HOT represents higher order terms of the expansion and can be neglected. Therefore, a linearized relationship between the deviation of the design-point to-be-adapted component parameters and the deviation of corresponding design-point target performance parameters of a gas turbine around an initial design point can be expressed with

$$\Delta\vec{z} = H \cdot \Delta\vec{x} \tag{3}$$

For the set of design-point to-be-adapted component parameters at initially selected design point, the corresponding predicted target performance parameters may be different from their actual values and such a difference is the deviation of the performance parameters relative to their target. For any expected deviation of engine target performance parameters $\Delta\vec{z}$ from its initial point, the deviation of to-be-adapted component parameters from its initial point can be predicted by inverting the influence coefficient matrix (ICM) H to a adaptation coefficient matrix (ACM) H^{-1} leading to Eq. (4) when $N=M$

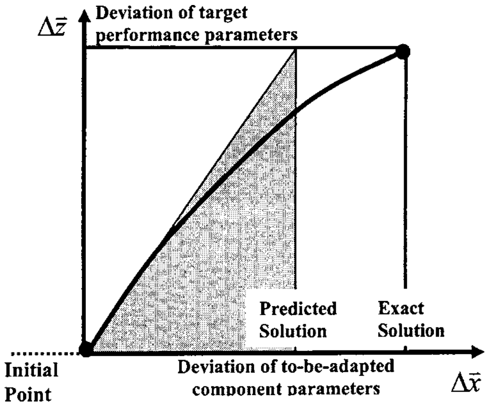


Fig. 1 Linear performance adaptation

$$\Delta\vec{x} = H^{-1} \cdot \Delta\vec{z} \tag{4}$$

The idea of the linear adaptation process is illustrated in Fig. 1, where the linear prediction of $\Delta\vec{x}$ will be a good estimate of the shift of engine to-be-adapted component parameters from their initial values when the shift is small and the thermodynamic behavior of the engine around the initial design point is close to linear.

It is very likely that when the nonlinearity of the relationship between the to-be-adapted component parameters and the target performance parameters is significant, the predicted deviation of to-be-adapted component parameters will be far from accurate. The nonlinearity of the engine thermodynamic behavior is taken into account by using the Newton-Raphson iterative method, Escher and Singh [8], where linear prediction process is applied iteratively until a converged solution is obtained (Fig. 2).

In many situations, the number of to-be-adapted component parameters N and the number of target performance parameters M may not be equal. If $N > M$, Eq. (3) is underdetermined. Typically, such a situation leads to an infinite number of least-squares solutions. A pseudoinverse is defined as

$$H^{\#} = H^T(HH^T)^{-1} \tag{5}$$

The solution resulting from the pseudoinverse

$$\vec{x} = H^{\#}\vec{z} \tag{6}$$

is best in a least-squares sense.

If $N < M$, Eq. (3) is overdetermined and there are redundant equations. A pseodoinverse is defined as

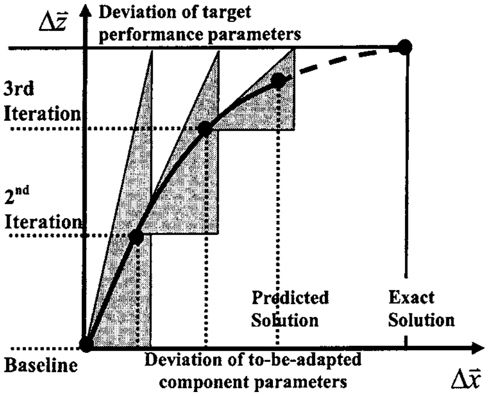


Fig. 2 Nonlinear performance adaptation

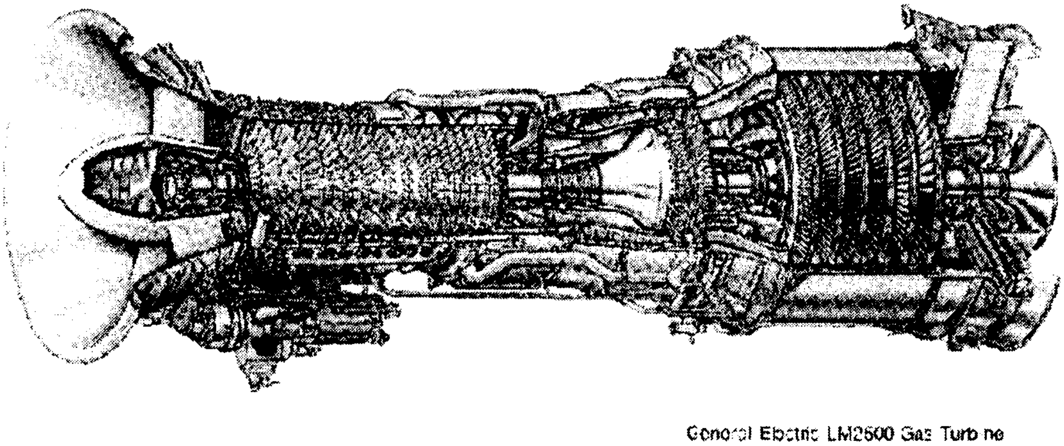


Fig. 3 GE LM2500+ gas turbine engine [10]

$$H^{\#} = (H^T H)^{-1} H^T \tag{7}$$

The resulting solution of Eq. (3), $\vec{x} = H^{\#}\vec{z}$, is also the best in a least-squares sense.

The convergence of the interative adaptation calculation process shown in Fig. 2 is declared when the predicted performance is very close to the real target performance. This criterion is shown in Eq. (8), where the root-mean square (rms) of the difference between predicted and target performance parameters is smaller than σ when convergence is declared

$$\mathrm{rms} = \sqrt{\frac{\sum_{i=1}^{M} (z_{i,\mathrm{predicted}} - z_{i,\mathrm{target}})^2}{M}} < \sigma \tag{8}$$

where σ is a very small number (0.001 is chosen in this study).

The adaptation error for each target performance parameter is described by the relative difference between the predicted and actual engine target performance parameter and expressed with

$$\varepsilon_i = \frac{z_{i,\mathrm{predicted}} - z_{i,\mathrm{target}}}{z_{i,\mathrm{target}}} 100\% \tag{9}$$

where ε_i is the adaptation error for ith target performance parameter.

The design-point gas turbine performance adaptation can also be seen as an optimization problem where the difference between the predicted and actual target performance parameters represented with rms in Eq. (8) can be used as an objective function for minimization and the to-be-adapted component parameters are searched with an optimization algorithm. Further discussion of this technique is beyond the scope of this paper, and therefore it is not discussed here.

LM2500+ and Performance Simulation

The gas turbine engine used in this research is a General Electric industrial LM2500+ installed in Isle of Man, UK by MANX Electricity Authority (MEA). MEA has two LM2500+ engines, each producing 29 MW (dry mode), driving a combined-cycle power plant with two once-through steam generators and a steam turbine producing 20–25 MW [9]. The LM2500+ industrial gas turbine (Fig. 3) is a two-shaft engine with one compressor, one burner, one compressor turbine, and one power turbine. The available nominal performance parameters of the engine (dry mode) are as follows [10]:

Exhaust gas flow rate	83 kg/s
Total pressure ratio	23.1
Power output	29 MW
Thermal efficiency	38.8%

Actual engine gas path performance parameters are measured regularly in MEA. One set of measurements were selected and used in this study. To use the measured data effectively, measurement uncertainty must be considered and proper data preprocessing must be used. The measurement uncertainties are normally due to the following reasons:

- engine normally runs at nonstandard ambient conditions
- measurement noise and bias
- nonuniformed radial and circumferential distribution of flow field even at same axial location of the engine
- the operating condition may be different from the nominal design-point condition

Therefore, the measurement chosen for the design-point performance adaptation was taken when the engine ran steadily at an operating point very close to the nominal design condition and regarded as design-point performance although it might be different from the design-point set by engine manufacturer. Necessary ambient and power condition corrections may be used if the measurements are made at a condition different from sea-level standard condition. Because of the above-mentioned reasons, the measured values of gas path parameters may not be the same as the actual average value of the parameters. Suppose that the measurement noise distribution is of Gaussian type, the distribution of the measurement of a gas path parameter would be that shown in Fig. 4 and the accuracy of the measurement is determined by the standard deviation σ and bias. Therefore, the values of the gas path parameters taken from the measurement can only be seen as an approximation of the true average values of those parameters. Therefore, the adapted engine component parameters corresponding to the measured performance parameters can only be seen as an approximation of their true values

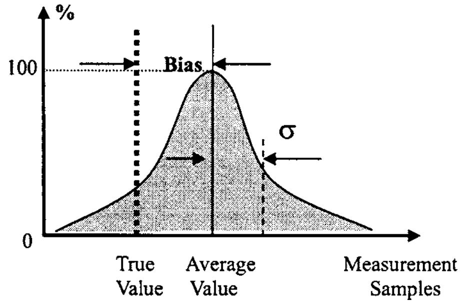


Fig. 4 Distribution of samples of a performance parameter measurement

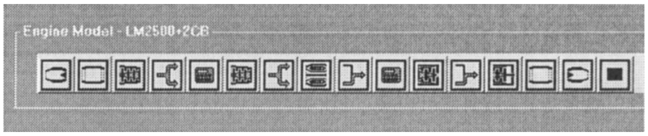


Fig. 5 Gas turbine model in PYTHIA

The adaptation starts from an initial design point obtained with initially guessed values of to-be-adapted component parameters. The initially obtained design-point performance of LM2500+ was simulated using software PYTHIA developed at the School of Engineering, Cranfield University. PYTHIA is gas turbine performance simulation and diagnostics software with a user friendly interface developed from TURBOMATCH [11]. A PYTHIA engine model is assembled with a collection of component bricks visually expressed with component icons, Fig. 5. The output of performance calculations are engine thrust or power, specific fuel consumption (SFC) or thermal efficiency, etc., together with details of individual component performance and of the gas properties at various gas path stations within the engine. The design-point performance adaptation described in this paper has been implemented recently.

Two different engine model configurations have been tested in the performance adaptation, one (model A) with two compressor bricks representing the compressor (Fig. 6) where the cooling flow for the compressor turbine is taken from the exit of the first compressor brick and the other (model B) with a single compressor brick for the same compressor (Fig. 7) where the cooling flow for the compressor turbine is taken from the exit of the compressor. Model A is more complicated than model B, but represents the real engine more realistically because the cooling flows for the turbines are taken from the right locations of the compressor, whereas model B is relatively simple.

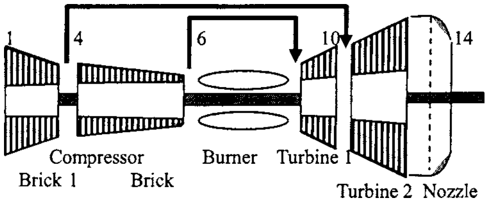


Fig. 6 LM2500+ engine model A

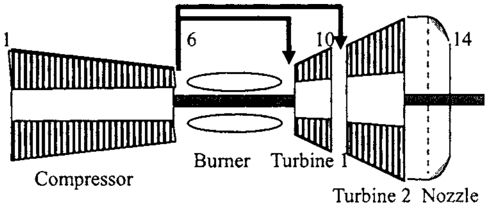


Fig. 7 LM2500+ engine model B

Table 2 Adaptation results with different model configurations (models A and B)

No.	To-be-adapted component parameter	Shifting from initial values after adaptation (%)	
		Model A	Model B
1	m_a	−7.833	−7.827
2	η_{c1}	3.835	−1.329
3	PR_1	fixed	−2.01
4	m_1	−34.9	19.34
5	m_2	0.23	0.866
6	PR_2	5.518	N/A
7	η_{c2}	0.294	N/A
8	ΔP_b	0.368	1.081
9	TET	−2.875	−0.172
10	η_{t1}	−4.513	3.303
11	η_{t2}	2.320	0.938

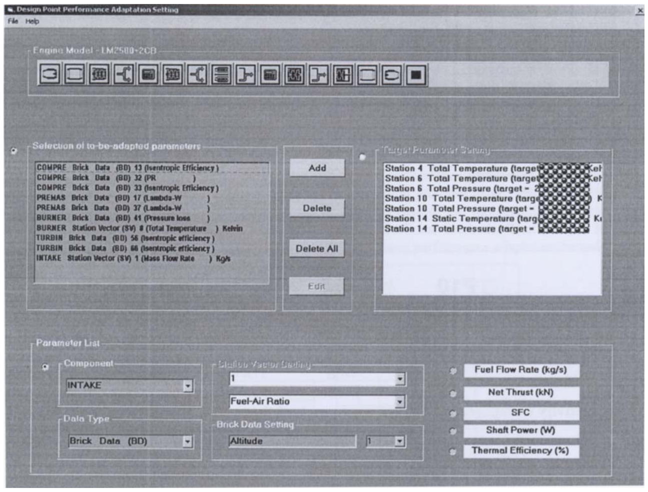
Selection of Adaptation Parameters and Sensitivity Analysis

Among the two types of adaptation parameters being used in the performance adaptation, the chosen target performance parameters at design-point are shown in Table 1 and the chosen to-be-adapted component parameters are shown in Table 2. The target performance parameters are total temperature for the cooling flow from midcompressor (T_4), total temperature (T_6) and total pressure (P_6) at compressor exit, total temperature (T_{10}) and total pressure (P_{10}) at compressor turbine exit, exhaust static temperature (Ts_{14}), engine thermal efficiency (η), and exhaust total pressure (P_{14}). The to-be-adapted component parameters are air mass flow rate (m_a), compressor pressure ratios ($\mathrm{PR}_1, \mathrm{PR}_2$), compressor isentropic efficiencies (η_{c1}, η_{c2}), cooling flows (m_1, m_2), burner pressure loss (ΔP_b), TET, compressor turbine efficiency (η_{t1}), and power turbine efficiency (η_{t2}); their nominal values may be available in an engine manual or the public domain, but their actual values may be slightly different even for engines of the same type due to manufacturing and assembly tolerance and engine health conditions. A set of chosen to-be-adapted engine component parameters and target performance parameters in PYTHIA is shown in Fig. 8.

Proper selection of the to-be-adapted component parameters is crucial to get converged and reasonable adaptation results. Some important factors should be taken into account in the selection of those to-be-adapted engine component parameters, as follows:

Table 1 Engine target performance parameters

No.	Target performance parameter	Model A		Model B	
		Initial error (%)	Adapting error (%)	Initial error (%)	Adapting error (%)
1	T_4	1.20	0	N/A	N/A
2	T_6	−0.32	-2×10^{-5}	−0.22	2×10^{-5}
3	P_6	−5.23	-3×10^{-5}	2.05	-8×10^{-6}
4	T_{10}	4.04	-1×10^{-5}	1.26	-1×10^{-5}
5	P_{10}	7.38	-1×10^{-4}	1.04	-4×10^{-5}
6	Ts_{14}	0.74	-7×10^{-5}	1.60	-2×10^{-4}
7	η	−10.4	8×10^{-5}	−10.33	-3×10^{-5}
8	P_{14}	22.21	-2×10^{-4}	17.82	5×10^{-5}



COMPRE Brick Data (BD) 13 (Isentropic Efficiency)
COMPRE Brick Data (BD) 32 (PR)
COMPRE Brick Data (BD) 33 (Isentropic Efficiency)
PREMAS Brick Data (BD) 17 (Lambda-W)
PREMAS Brick Data (BD) 37 (Lambda-W)
BURNER Brick Data (BD) 41 (Pressure loss)
BURNER Station Vector (SV) 8 (Total Temperature) Kelvin
TURBIN Brick Data (BD) 56 (Isentropic efficiency)
TURBIN Brick Data (BD) 66 (Isentropic efficiency)
INTAKE Station Vector (SV) 1 (Mass Flow Rate) Kg/s

Station 4 Total Temperature (target [redacted] Kel
Station 6 Total Temperature (target Kel
Station 6 Total Pressure (target = 2
Station 10 Total Temperature (targe) K
Station 10 Total Pressure (target =
Station 14 Static Temperature (targ K
Station 14 Total Pressure (target =

Component: INTAKE
Station Vector Setting: 1
Fuel-Air Ratio
Data Type: Brick Data (BD)
Brick Data Setting: Altitude 1

Fig. 8 Setting of adaptation parameters in PYTHIA

- They should have functional relationship with the target performance parameters. Otherwise, they do not provide more dimensions in adaptation and also do not contribute to better adaptation results.
- The number of to-be-adapted component parameters should be at least equal to the number of the target performance parameters. More to-be-adapted component parameters will provide more freedom in adaptation and be easier to converge but also likely to produce multiple solutions.

The functional relationship between the target performance parameters and the to-be-adapted component parameters can be assessed with a sensitivity analysis shown in Figs. 9(*a*) and 9(*b*). It can be seen that the target performance parameters are influenced differently by the to-be-adapted component parameters and the levels of influence can be significantly different. For example, T_4 is only influenced slightly by η_{c1}, and P_{14} is influenced significantly by turbine entry temperature (TET) but also influenced effectively by m_a, η_{c1}, η_{c2}, η_{t1}, and η_{t2}.

Initial values for the to-be-adapted component parameters must be given to obtain the starting point for the performance adaptation. Because of the iterative nature of the nonlinear adaptation method used in this study to take into account the nonlinearity of the thermodynamic relationship within the engine, the initial values chosen for the to-be-adapted component parameters close to the solution will enhance the convergence of the adaptation. Therefore, initial manual selection of the to-be-adapted component parameters should be done in such a way that a resulted set of target performance parameters are as close to their actual values as possible. Then the more difficult final adaptation of the performance to the target performance is left to the design-point performance adaptation technique. Based on authors' experience, around 5% of adaptation shift can be accommodated by the adaptation algorithm.

Adaptation With Different Engine Model Configurations

For the same engine, different engine model configurations may result in very different adapted engine component parameters.

The adaptation approach is applied to two slightly different model engine configurations: Model A (Fig. 6) and model B (Fig. 7). The adapted target performance errors and corresponding to-be-adapted component parameters are shown in Tables 1 and 2.

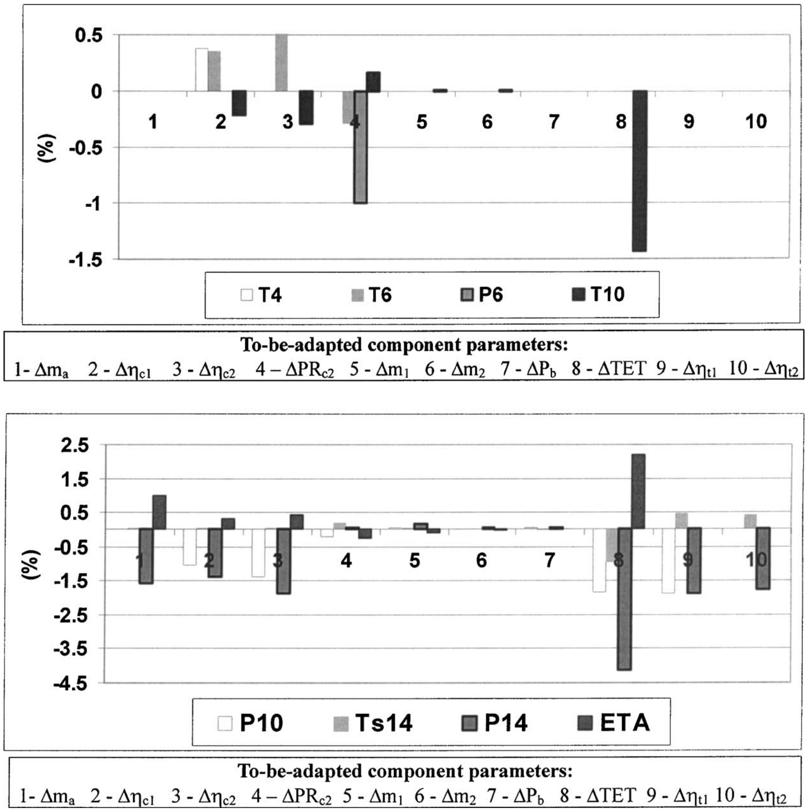
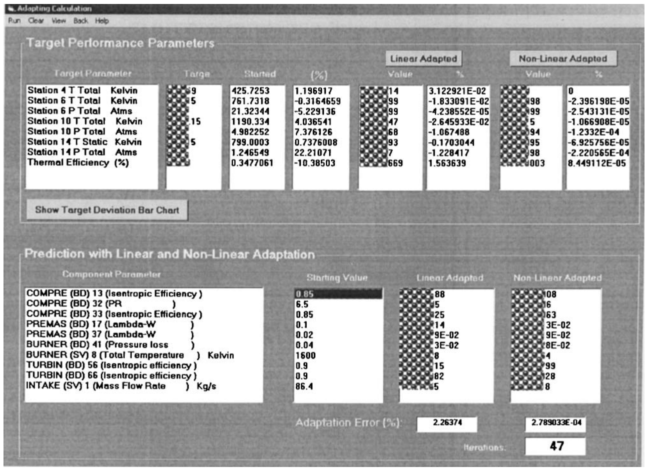


Fig. 9 (*a*) Sensitivity comparison (−1% change to each of the to-be-adapted component parameters) and (*b*) sensitivity comparison (−1% change to each of the to-be-adapted component parameters)



Target Performance Parameters

Target Parameter	Targe	Started	(%)	Linear Adapted Value	%	Non-Linear Adapted Value	%
Station 4 T Total Kelvin	[redacted]	425.7253	1.196917	[redacted]	3.122921E-02	[redacted]	0
Station 6 T Total Kelvin	[redacted]	761.7318	-0.3164659	[redacted]	-1.833091E-02	[redacted]	-2.396198E-05
Station 6 P Total Atms	[redacted]	21.32344	-5.229136	[redacted]	-4.238552E-05	[redacted]	-2.543131E-05
Station 10 T Total Kelvin	[redacted]	1190.334	4.036541	[redacted]	-2.645933E-02	[redacted]	-1.066908E-05
Station 10 P Total Atms	[redacted]	4.982252	7.376126	[redacted]	-1.067488	[redacted]	-1.2332E-04
Station 14 T Static Kelvin	[redacted]	799.0003	0.7376008	[redacted]	-0.1703044	[redacted]	-6.925756E-05
Station 14 P Total Atms	[redacted]	1.246549	22.21071	[redacted]	-1.228417	[redacted]	-2.220565E-04
Thermal Efficiency (%)	[redacted]	0.3477061	-10.38503	[redacted]	1.563639	[redacted]	8.449112E-05

Prediction with Linear and Non-Linear Adaptation

Component Parameter	Starting Value	Linear Adapted	Non-Linear Adapted
COMPRE (BD) 13 (Isentropic Efficiency)	0.85	[redacted]	[redacted]
COMPRE (BD) 32 (PR)	6.5	[redacted]	[redacted]
COMPRE (BD) 33 (Isentropic Efficiency)	0.85	[redacted]	[redacted]
PREMAS (BD) 17 (Lambda-W)	0.1	[redacted]	[redacted]
PREMAS (BD) 37 (Lambda-W)	0.02	[redacted]	[redacted]
BURNER (BD) 41 (Pressure loss)	0.04	[redacted]	[redacted]
BURNER (SV) 8 (Total Temperature) Kelvin	1600	[redacted]	[redacted]
TURBIN (BD) 56 (Isentropic efficiency)	0.9	[redacted]	[redacted]
TURBIN (BD) 66 (Isentropic efficiency)	0.9	[redacted]	[redacted]
INTAKE (SV) 1 (Mass Flow Rate) Kg/s	86.4	[redacted]	[redacted]

Adaptation Error (%): 2.26374 | 2.789033E-04

Iterations: 47

Fig. 10 Adaptation result shown in PYTHIA

Table 4 Adaptation results with linear and nonlinear adaptations

No.	To-be-adapted component parameter	Shifting from initial values after adaptation (%)	
		Linear adaptation	Nonlinear adaptation
1	m_a	–9.40	–7.833
2	η_{c1}	3.732	3.835
3	PR_1	fixed	fixed
4	m_1	–37.18	–34.9
5	m_2	–0.336	0.23
6	PR_2	5.52	5.518
7	η_{c2}	0.404	0.294
8	ΔP_b	0.268	0.368
9	TET	–2.917	–2.875
10	η_{t1}	–5.114	–4.513
11	η_{t2}	3.876	2.320

The shifting of the to-be-adapted parameters from their initial values for the two engine model configurations shown in Table 2 are different, which shows that the adaptation provides different to-be-adapted component parameters for different model configurations. In other words, the representation of the engine with different model configurations results in different adaptation results, although both of the adapted engine models can match the target performance very well (Table 1). The adaptation result with model A provided by PYTHIA is shown in Fig. 10.

Comparison Between Linear and Nonlinear Adaptation Methods

In theory, the nonlinear adaptation technique should provide more accurate adapted results than the linear partner as the former takes into account the nonlinearity of the thermodynamic relationship within the engine. This is proved in Tables 3 and 4 where model A configuration is used in the adaptation and the linear and nonlinear adaptations are applied to the same set of to-be-adapted component parameters and target performance parameters. It is obvious that the nonlinear adaptation provides better adapted performance than the linear one because of the fundamental difference between the linear and nonlinear methods. Therefore, it is suggested that only nonlinear performance adaptation be used in applications.

Robustness and Computation Speed

In the performance adaptation with Model A, the number of to-be-adapted parameters is 10 while the number of target performance parameters is 8. Changing the number of to-be-adapted component parameters and target performance parameters is likely to get different solutions. One example of this is shown in Tables 5 and 6, where one of the target performance parameters is excluded from the list in case 2 while other settings are the same. The nonlinear adaptation process is applied to the two cases, and two adaptation results are obtained. It can be seen in Table 5 that all the target performance parameters in case 2 are satisfied with small adaptation errors similar to case 1, but the obtained engine to-be-adapted component parameters are very different to case 1, which can be seen from the different values of the shifting from their initial values in Table 6.

Table 3 Comparison of linear and nonlinear adaptations

No.	Target performance parameter	Initial error (%)	Linear adaptation error (%)	Nonlinear adaptation error (%)
1	T_4	1.20	3×10^{-2}	0
2	T_6	–0.32	-2×10^{-2}	-2×10^{-5}
3	P_6	–5.23	-4×10^{-5}	-3×10^{-5}
4	T_{10}	4.04	-3×10^{-2}	-1×10^{-5}
5	P_{10}	7.38	–1.07	-1×10^{-4}
6	Ts_{14}	0.74	–0.17	-7×10^{-5}
7	η	–10.4	1.56	8×10^{-5}
8	P_{14}	22.21	–1.23	-2×10^{-4}

Different selection of the to-be-adapted component parameters may have significant effects on the performance adaptation. Exclusion of cooling flow (m_2) and burner pressure loss (ΔP_b) do not affect the result as they have little influence on the model engine performance, which can also be seen in the sensitivity analysis shown in Figures 9(*a*) and 9(*b*). Exclusion of any sensitive to-be-adapted component parameters may result in the divergence of the iteration in the nonlinear adaptation. This means that the adaptation process must include the most important to-be-adapted component parameters. If they are not included, reasonable adapted engine model may not be obtained and the adaptation process may diverge. The sensitivity of to-be-adapted component parameters to engine target performance parameters can be assessed through the sensitivity analysis described in a previous section.

The computing speed for both the linear and nonlinear adaptation calculation is very fast. It takes a fraction of second for linear adaptation calculation and a few seconds for nonlinear adaptation calculation using a personal computer with Pentium IV processor.

Conclusions

In this paper, a detailed description of a novel gas turbine design-point performance adaptation approach has been given. Application of the approach to the real performance matching problem of an industrial gas turbine engine shows that:

- The approach is effective in producing a good match of an engine model to measured or target engine performance.
- The approach is fast in computation.
- Different engine model configurations may have significant influence on the adapted results.
- Different selection of to-be-adapted component parameters and target performance parameters may produce different adaptation results.
- Improper selection of to-be-adapted component parameters and target performance parameters may result in the divergence of adaptation calculation.
- Nonlinear adaptation approach can produce much better

Table 5 Multiple adaptation results-target performance parameters

		Case 1		Case 2	
No.	Target performance parameter	Initial error (%)	Adapting error (%)	Initial error (%)	Adapting error (%)
1	T_4	1.20	0	1.20	7×10^{-6}
2	T_6	−0.32	-2×10^{-5}	−0.32	6×10^{-5}
3	P_6	−5.23	-3×10^{-5}	−5.23	-4×10^{-5}
4	T_{10}	4.04	-1×10^{-5}	4.04	-6×10^{-5}
5	P_{10}	7.38	-1×10^{-4}	7.38	-5×10^{-4}
6	Ts_{14}	0.74	-7×10^{-5}	0.74	2×10^{-4}
7	η	−10.4	8×10^{-5}	excluded	
8	P_{14}	22.21	-2×10^{-4}	22.21	-8×10^{-4}

Table 6 Multiple adaptation results-To-be-adapted component parameters

		Shifting from initial values after adaptation (%)	
No.	To-be-adapted component parameter	Case 1	Case 2
1	m_a	−7.833	−0.174
2	η_{c1}	3.835	3.835
3	PR_1	fixed	fixed
4	m_1	−34.9	−0.475
5	m_2	0.23	0.025
6	PR_2	5.518	5.518
7	η_{c2}	0.294	0.294
8	ΔP_b	0.368	0.094
9	TET	−2.875	−2.615
10	η_{t1}	−4.513	−3.761
11	η_{t2}	2.320	−3.431

adapted engine model than its linear partner and, therefore, only the nonlinear adaptation approach is recommended for applications.

- The adapted engine performance model can only be seen as a better approximation of real engine performance at its design-point. The predicted values of to-be-adapted component parameters may be different from those of real engines and the accuracy of the adapted engine model is determined by the amount of available target performance parameters and their accuracy.

Acknowledgment

The present research is supported by a Platform Grant of the Engineering and Physical Sciences Research Council (EPSRC), United Kingdom. The authors also thank the MANX Electricity Authority, Isle of Man, UK, for allowing us to use their test data and providing valuable technical information.

Nomenclature

H = influence coefficient matrix (ICM)
H^{-1} = inverse of H/adaptation coefficient matrix (ACM)
$H^{\#}$ = pseudoinverse matrix of H
HOT = high-order terms
$h(\,)$ = A function
M = dimension of target performance parameter vector
m_a = engine inlet air flow rate (kg/s)
m_1/m_2 = cooling flow rate to turbine 1 and 2
N = dimension of to-be-adapted component parameter vector
P = total pressure (atm)
$\mathrm{PR}_{c1}/\mathrm{PR}_{c2}$ = pressure ratio of compressor brick 1 and 2
rms = root mean square
SFC = specific fuel consumption (mg/N s)
T = total temperature (K)
TET = turbine entry temperature (K)
Ts = static temperature (K)
$\vec{x}$ = to-be-adapted component parameter vector
$\vec{z}$ = target performance parameter vector
ε = adaptation error
Σ = summation
η (ETA) = engine thermal efficiency
η_{c1} = isentropic efficiency of compressor 1
η_{c2} = isentropic efficiency of compressor 2
η_{t1} = isentropic efficiency of turbine 1
η_{t2} = isentropic efficiency of turbine 2
ΔP_b = burner relative pressure loss

Subscripts

b = burner
0 = starting point/baseline
4 = compressor 1 exit
6 = burner inlet
10 = compressor turbine (turbine 1) exit
14 = nozzle exit

References

[1] Stamatis, A., Mathioudakis, K., and Papailiou, K. D., 1990, "Adaptive Simulation of Gas Turbine Performance," ASME J. Eng. Gas Turbines Power, **112**(2), pp. 168–175.
[2] Lambiris, B., Mathioudakis, K., and Papailiou, K. D., 1991, "Adaptive Modeling of Jet Engine Performance With Application to Condition Monitoring," ISABE Paper No. 91-7058.
[3] Stamatis, A., Mathioudakis, K., and Papailiou, K., 1992, "Optimal Measurement and Health Index Selection for Gas Turbine Performance Status and Fault Diagnosis," ASME J. Eng. Gas Turbines Power, **114**, pp. 209–216.
[4] Kong, C., Ki, J., and Kang, M., 2002, "A New Scaling Method for Component Maps of Gas Turbine Using System Identification," ASME Paper No. GT-2002-30150.
[5] Kong, C., Kho, S., and Ki, J., 2004, "Component Map Generation for a Gas Turbine Using Genetic Algorithms," ASME Turbo Expo 2004, June, ASME Paper No. GT2004-53736.
[6] Roth, B. A., Mavris, D. N., and Doel, D. L., 2003, "Estimation of Turbofan Engine Performance Model Accuracy and Confidence Bounds," 16th International Symposium on Air Breathing Engines, ISABE Paper No. 2003-1208.
[7] Roth, B. R., Mavris, D., Doel, D. L., and Beeson, D., 2004, "High-Accuracy Matching of Engine Performance Models to Test Data," ASME Turbo Expo 2004, June, ASME Paper No. GT2003-38784.
[8] Escher, P. C., and Singh, R., 1995, "An Objective-Oriented Diagnostics Computer Program Suitable for Industrial Gas Turbines," 21st (CIMAC) International Congress of Combustion Engines, Switzerland, May 15–18.
[9] Jeffs, E., 2004, "Isle of Man Plant Has Novel Design," Turbomachinery International Magazine, **45**(3), pp. 27–28.
[10] GE Leaflet "GE Marine & Industrial Engines: LM2500+ Gas Turbine" or http://www.gepower.com.
[11] Macmillan, W. L., 1974, "Development of a Module Type Computer Program for the Calculation of Gas Turbine Off Design Performance," Ph.D. thesis, Cranfield University, England.

A Sequential Approach for Gas Turbine Power Plant Preventative Maintenance Scheduling

Yongjun Zhao

Vitali Volovoi

Mark Waters

Dimitri Mavris

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

Traditionally, gas turbine power plant preventive maintenance schedules are set with constant intervals based on recommendations from the equipment suppliers. Preventive maintenance is based on fleet-wide experience as a guideline as long as individual unit experience is not available. In reality, the operating conditions for each gas turbine may vary from site to site and from unit to unit. Furthermore, the gas turbine is a repairable deteriorating system, and preventive maintenance usually restores only part of its performance. This suggests a gas turbine needs more frequent inspection and maintenance as it ages. A unit-specific sequential preventive maintenance approach is therefore needed for gas turbine power plant preventive maintenance scheduling. Traditionally, the optimization criteria for preventive maintenance scheduling is usually cost based. However, in the deregulated electric power market, a profit-based optimization approach is expected to be more effective than the cost-based approach. In such an approach, power plant performance, reliability, and the market dynamics are considered in a joint fashion. In this paper, a novel idea that economic factors drive maintenance frequency and expense to more frequent repairs and greater expense as equipment ages is introduced, and a profit-based unit-specific sequential preventive maintenance scheduling methodology is developed. To demonstrate the feasibility of the proposed approach, a conceptual level study is performed using a base load combined cycle power plant with a single gas turbine unit. [DOI: 10.1115/1.2179470]

Gas Turbine Power Plant Preventive Maintenance Scheduling

The deregulation of the electric power market has introduced a strong element of competition. As a result, power plant operators are striving to develop advanced operational strategies to maximize the profitability in the dynamic electric power market.

A systematic approach for profit-based outage planning is introduced in Ref. [1], with consideration given to system performance, the aging and reliability of equipment, maintenance practices, and market dynamics accounting for the price and availability of fuel as well as the generation of revenues in competing markets. A dual time-scale method is developed to project coupled optimal generation scheduling and outage planning for a single operations and maintenance cycle. This paper studies gas turbine power plant maintenance scheduling with consideration given to multiple operations and maintenance cycles. Specifically, the impact of unit aging on maintenance frequency is investigated over the life cycle of power plants.

Gas turbine units are widely used for land electric power generation, and maintenance planning has a strong impact on the profitability of a gas turbine power plant. Performance requirements for modern heavy-duty gas turbines necessitate extreme operating conditions for hot gas path components. As a result, these critical components have a limited life span and, more generally, a gas turbine represents an *aging* system experiencing continuous degradation during its operation. This physical degradation manifests itself in performance degradation, as well as in an increased risk of forced outage. Operating conditions of gas turbines determine the aging processes (and degradation rates) of their components and, therefore, affect both reliability and performance degradation. Timely preventive maintenance is scheduled to stop further degradation and to partially restore performance and reliability.

Maintenance scheduling is an important task for power plant operation. The maintenance can be defined as the combination of all actions intended to maintain the plant or to restore it to a performance level in terms of power output and heat rate so that the plant can perform its required functions economically [2]. Maintenance activities include inspection, repair, and replacement, and they constitute a significant proportion of the variable operating cost. The types of maintenance inspections for gas turbines can be classified as standby, running, and disassembly inspections. Disassembly inspection is one that requires opening the turbine to inspect internal components, and it can be further classified as combustion inspection, hot gas path inspection, and major inspection [3]. The effective scheduling of these disassembly inspection/maintenance actions is the primary interest of this study.

The decisions to be made for the maintenance inspection problem should address two issues: when the next inspection should occur (i.e., timing) and what maintenance action to take. In this study, the emphasis is on the determination of the optimal timing of preventive maintenance.

Preventive Maintenance Models

Maintenance can be classified into two major categories: corrective and preventive. There are two commonly used preventive maintenance policies: periodic preventive maintenance and sequential preventive maintenance. Under a periodic preventive maintenance policy, a system is maintained at integer multipliers of some fixed time period. Under sequential preventive maintenance, the system is maintained at a sequence of intervals that may have unequal lengths of intervals [4], determined by the actual status of operating parameters.

In the last several decades, numerous models for optimally scheduling inspections and/or maintenance have been published in the literature [4–11]. The periodic preventive maintenance policy

---

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the Journal of Engineering for Gas Turbines and Power. Manuscript received February 18, 2005; final manuscript received December 14, 2005. Review conducted by M. Boyce.

has been extensively used, and one of the reasons for this is that maintenance is easy to schedule. However, the sequential preventive maintenance policy is more realistic in that most systems need more frequent maintenance as they age, reflecting the fact that usually preventive maintenance does not fully restore the power plant to its original "as good as new" state (see below).

Early studies of maintenance models usually assumed that, after corrective or preventive maintenance, the system is in one of the two extreme conditions, either as good as new or as bad as old (the latter type of maintenance restores the system to the state immediately preceding the failure, see below). Furthermore, downtime due to maintenance is considered negligible and thus discounted, and the aging of the unit is not considered [4–11]. For real systems such as gas turbine power plants, these assumptions are not true.

Most preventive maintenance improves or restores the system, but the improvement depends on the age of the system as well as the cost and time of the preventive maintenance actions [10]. The effect of maintenance usually is somewhere between as good as new and as bad as old. Therefore, most systems need more frequent maintenance due to aging and imperfect maintenance [9].

Reviews and surveys of preventive maintenance models for deteriorating single-unit system have been published in the literature [5,8]. A classification of the maintenance practices based on the maintenance effectiveness is introduced in Ref. [12]. Five categories, according to the degree to which the operating conditions of an item is restored by maintenance, are identified, and they are *perfect, minimal, imperfect, worse,* and *worst* maintenance. The perfect repair or perfect maintenance is an action that restores the system to as good as new. The system has the same reliability distribution as a brand new one after perfect maintenance. Minimal repair or minimal maintenance is an action that restores the system to the as-bad-as-old state with the failure rate unchanged from the moment when the system failed. Imperfect repair or imperfect maintenance is an action that restores the system operating state to somewhere between as good as new and as bad as old [12].

Minimal repair is a frequently used assumption in the literature [11]. This assumption is acceptable for a complex system with many components, and the failure of each component will lead to the failure of the entire system. The operating status of the whole system will not change much if one or some of its components are replaced or repaired, since it has so many components [13]. It is assumed that in this study that gas turbine power plants are complex systems with numerous components and that the corrective maintenance of a gas turbine power plant is minimal maintenance.

The preventive maintenance for a gas turbine power plant includes combustion inspection, hot gas path inspection, and major inspection, and each can be classified as a different type of maintenance. In some references, these maintenance actions are referred to as overhauls. The overhaul is scheduled and may act on groups of components, and therefore, they can be more effective on the restoration of a system's performance and reliability than would minimal maintenance [11]. For example, the combustion inspection is inspection that concentrates on the combustion liners, transition pieces, fuel nozzles, and end caps. Hot gas path inspection is an inspection performed to inspect those parts exposed to hot gas discharged from the combustion process, and it includes the full scope of combustion inspection and a detailed inspection of turbine nozzles, stationary stator shrouds, and turbine buckets. Major inspection is a more extensive inspection, which includes the work scope of both combustion and hot gas path inspections, and it also examines all of the major flange-to-flange components of the gas turbine, which are subject to deterioration during normal turbine operation [3].

Although the major preventive maintenance actions can rejuvenate the gas turbine power plant system, they cannot restore it to as-good-as-new state, as they do not eliminate all the performance and reliability degradation that has taken place in this complex system. As a result, the major preventive maintenance will restore the gas turbine power plant to be somewhere between as good as new and as bad as old. This is referred to as imperfect maintenance.

The Need for Sequential Preventive Maintenance Scheduling

Traditionally, gas turbine power plant preventive maintenance is scheduled with constant maintenance intervals (factored starts and operating hours) based on recommendations from the equipment suppliers. The preventive maintenance actions are based on fleet-wide experiences and are scheduled in a one-size-fits-all fashion. This constant maintenance interval philosophy is referred to as periodic maintenance and is not able to take into account the gas turbine system as a repairable aging system.

However, the gas turbine is an aging system and the aging of the power plant heavily depends on the operating conditions. In reality, the operating conditions of gas turbine power plants vary from site to site and unit to unit. Maintenance performed without regard to the condition of the equipment may result in wasted resources for equipment that is aging less rapidly than expected or equipment may experience high risk of failure if the equipment ages more rapidly than expected. This suggests that each unit should be individually treated. For such a unit-specific approach to be successful, accurate predictions of reliability and performance degradation for each gas turbine is necessary.

Performance and reliability degradation increase as the power plant ages, and, as discussed above, the major maintenance action will partially restore performance and improve reliability. As a result, the gas turbine power plant is an aging system in that, for any age x, its failure rate $h_{m+1}(x)$ during the $(m+1)^{\text{th}}$ operations and maintenance cycle is strictly larger than $h_m(x)$ during the m^{th} operations and maintenance cycle[1] (O&M cycle hereafter). Here x is the elapsed age from the end of each major preventive maintenance action.

To consider the aging of the gas turbine power plant, a sequential preventive maintenance philosophy is needed. Here the maintenance intervals are subject to change, and the length of the maintenance intervals are determined by the combination of performance degradation, reliability, and market signals.

Optimization Criteria

The first task of preventive maintenance scheduling is to set the optimization objective. Traditionally, for complicated systems (such as a gas turbine power plant), maintenance cost and availability are two of the most important concerns to the equipment owner. Several optimization criteria that have been published in the literature [5,6] are as follows:

- minimize system maintenance cost rate
- maximize system reliability/availability
- minimize system maintenance cost rate while the system meets its reliability requirement
- maximize system reliability while the system meets its maintenance cost requirement

Many cost-based inspection and preventive maintenance policies have been published in the literature. For the cost-based maintenance scheduling, the optimization criteria is usually to minimize the long-run expected cost per unit time (the expected cost rate). However, in the deregulated electric power market, cost and reliability are not the only concerns: the ultimate goal of the

---

[1] An operations and maintenance cycle is the time period that includes a major maintenance (including combustion inspection, hot gas path inspection, or major inspection), and the subsequent continuous operating period.

power plant operator is to make a profit. Furthermore, in a market-based environment, the electricity market shows strong dynamics, and an optimized maintenance cost and maximized plant availability does not necessarily imply optimized profitability, since other factors, such as fuel cost, electricity price, and power demand and supply, also play a big role. This suggests that in a market-based environment, maintenance practices should be optimized to achieve maximum profit. Any optimization criteria aiming to maximize system availability and/or minimize cost only will inevitably lead to suboptimal solutions.

Constructing a profit function that incorporates availability and cost functions along with the revenue gained per unit operating time is described in Ref. [14]. However, the profit-based approach requires a great deal of information. Although profit-based unit commitment methodology has been published [15], profit based preventive maintenance scheduling has rarely been seen in the literature. A framework for a profit based, life-cycle oriented, and unit-specific operational optimization for a gas-turbine-based power plant is introduced in Ref. [16], and an implementation of profit-based outage planning coupled with generation scheduling is introduced in Ref. [1]. Also in Ref. [1], a combined consideration of performance degradation, loss in reliability, and market signals is presented. For the unit-specific maintenance approach, accurate performance degradation and reliability distribution for each gas turbine power plant is necessary, which requires realistic performance and reliability modeling based on unit operating conditions and maintenance history.

Modeling of Power Plant Operations and Maintenance

In the integrated framework introduced in Ref. [16], the power pant performance, reliability, and market dynamics are considered in an integrated fashion. This method is applicable to different categories of operational optimization problems and is employed here for the modeling of power plant operations and maintenance. A brief summary to the method is given in the section below. The key elements that define the power plant profit is the value of power or gross revenue due to the selling and cost of electricity, which includes cost of fuel, cost of operations and maintenance (excluding cost of fuel), and depreciation of the power plant investment. The following factors are pertinent to the cost and revenue of power plant operations.

The Electricity Market and Weather Conditions. In a market-based operation environment, price of electricity and price of fuel are major driving factors for power plant operational planning. The weather conditions, i.e., the ambient temperature, ambient pressure, and relative humidity, are also important factors, which show strong seasonal and daily trends and are stochastic in nature. All of these factors are important, and they simultaneously impact both gas turbine performance and the demand for electric power. In this study, three major external factors are investigated: price of electricity, price of fuel, and ambient temperature. Let t be the calendar time. The price of electricity, $M_p(t)$, price of fuel $F_c(t)$, and ambient temperature $T_a(t)$, are all functions of calendar time. It is understood that a relatively simplistic dynamic model employed here captures the essential dynamics expressed as daily variance, seasonal trend, and long-term trend.

Evaluation of Aging. The evaluation of the aging of a gas turbine power plant is an important task for determining power plant performance and reliability degradation. It is assumed that the system ages only when it is in operation and as it accumulates operating hours. One approach to define the age of the gas turbine power plant is the independent starts and hours (ISH) method. To define the age of a gas turbine, the knowledge of both factored fired hours and factored starts are needed. In the independent starts and hours approach, the age of the gas turbine L_{ISH}, using the independent starts and hours criteria, is given by

$$L_{\text{ISH}} = (H_f, S_f) \tag{1}$$

where H_f is the factored fired hours and S_f is the factored fired starts.

A baseline condition for operating hours is defined as a gas fuel unit operating at continuous duty, and with no water or steam injection [3]. The maintenance factor for this baseline is defined as 1.0. For operation that differs from the baseline, maintenance factors reflect the increased level of maintenance that is required. In so doing, the influence of factors, such as fuel type and quality, firing temperature setting, and the amount of water or steam injection, are considered with regard to an hours-based criterion. Similarly, a baseline condition for starts and maintenance factors can be defined based on the attributes of an actual start. Startup rate and the number of trips are considered for a start-based criterion [3].

Let $m_h(t)$ be the maintenance factor of operating hours at time t, and $m_s(i)$ the maintenance factor of start i. The factored fired hours H_f for time period T is defined below

$$H_f = \int_T m_h(t)dt \tag{2}$$

For starts, assume there are N startups during the time period T. The factored starts are defined as

$$S_f = \sum_{i=1}^{N} m_s(i) \tag{3}$$

Power Plant Performance. Power plant performance (output power rate and heat rate) is a function of power plant design, technology upgrades, operating mode, ambient conditions, and degree of degradation. Let $P(t)$ be the power output rate of the power plant and $\text{HR}(t)$ the heat rate. Power output rate and heat rate are given by Eqs. (4) and (5), respectively.

$$P(t) = P\begin{pmatrix}\text{system design, technology upgrade,} \\ \text{O}_\text{p}(t), \text{T}_a(t), \text{degradation}\end{pmatrix} \tag{4}$$

$$\text{HR}(t) = \text{HR}\begin{pmatrix}\text{system design, technology upgrade,} \\ \text{O}_\text{p}(t), \text{T}_a(t), \text{degradation}\end{pmatrix} \tag{5}$$

Here, O_P is operating profile type and $T_a(t)$ are ambient conditions.

The actual output rate and heat rate of the power plant must include the effects of the degradation. Let P_0 and HR_0 be the output rate and heat rate of the power plant at the beginning of its service life when no performance degradation has occurred. Let $\Delta^P(t)$ be the degradation of output rate at time t and $\Delta^{\text{HR}}(t)$ be the degradation of heat rate at time t. The degraded output rate and heat rate can be calculated using the following:

$$P(t) = P_0[1 - \Delta^p(t)] \tag{6}$$

$$\text{HR}(t) = \text{HR}_0[1 - \Delta^{\text{HR}}(t)] \tag{7}$$

Performance degradation is a function of system design and unit usage history, with the latter including both unit operating history and maintenance activities. In this study, the assumed shape of the typical performance deterioration versus service time curve, introduced by Diakunchak [17], is employed. Historical operational data are used to perform the performance degradation regression analysis. In this study, performance degradation of the power plant is modeled as a function of its actual operating hours. Let τ here be the age of the power plant. Assume the engine is with age τ_0 when it enters its service and assume the equation that defines the curve is in the form given by

$$\Delta(\tau) = a \ln[b(\tau + \tau_0) + c] + \Delta_0 \tag{8}$$

where Δ is the percentage of performance loss, which includes both power output and heat rate. a, b, c, and Δ_0 are parameters that depend on the configuration and usage history of the engine.

Maintenance Effectiveness. The primary maintenance means for performance restoration include online water wash, offline water wash, and major preventive maintenance actions, including combustion inspection, hot gas path inspection, and major inspection. These maintenance practices, such as water wash or hot gas path parts replacement, will restore part of the performance, which improves the status of the engine. The imperfect maintenance model is employed to study maintenance effectiveness.

Power plant performance and reliability can be modeled as functions of some measure of age. In this study, the performance degradation is modeled as a function of cumulative fired hours, and reliability is modeled as a function of plant factored fired hours and starts.

Hence, an approach to model maintenance effectiveness is to model the impact of maintenance on plant age. A virtual age method is introduced by Kijima et al. to model imperfect maintenance [18]. Let τ_m be the actual age of the item when it undertakes the m^{th} maintenance action. Let ε_m be the age reduction factor due to the m^{th} maintenance action, V_m^- be the virtual age of the component immediately before it undertakes the m^{th} maintenance action, and V_m^+ be the virtual age of the component immediately after it undertakes the m^{th} maintenance action. The age of the item right after it undertakes the first maintenance action is

$$V_m^+ = V_{m-1}^- + \varepsilon(\tau_m - \tau_{m-1}) \tag{9}$$

The impact of maintenance on performance and reliability can then be evaluated by the evaluating the impact of maintenance on the virtual age of the plant.

Cost of Maintenance. Assume a preventive maintenance is performed with cost C_{pm} as scheduled if the gas turbine does not fail during the given time period of operation, in which case corrective maintenance would be performed whenever the system fails.The maintenance cost here includes cost of preventive maintenance and cost of failure, which, in turn, includes the cost of corrective maintenance and loss of revenue due to unavailability of the plant. For a given time period T, the maintenance cost is therefore the summation of the cost of preventive maintenance and the cost of failure, which is given by

$$C_{om}(T) = C_{pm}(T) + C_{\text{failure}}(T) \tag{10}$$

The cost of failure C_{failure} can be given as

$$C_{\text{failure}} = C_{cm} + C_{lr} \tag{11}$$

where C_{cm} is the cost of corrective maintenance and C_{lr} is the cost due to a loss of revenue.

The expected cost of preventive maintenance during time period T is

$$E[C_{pm}(T)] = C_{pm}\int_{\tau_{pm}}^{\infty} f(t)dt \tag{12}$$

Assume the reliability of the investigated item is a three-parameter Weibull distribution. Then,

$$f(t) = \frac{\beta(t-t_0)^{(\beta-1)}}{\eta^{\beta}} e^{-[(t-t_0)/\eta]^{\beta}} \tag{13}$$

where $\eta > 0$ is the scale parameter, $\beta > 0$ is the shape parameter, and t_0 is the location parameter.

The expected cost of maintenance is given by

$$E[C_{om}(T)] = C_{pm}\int_{\tau_{pm}}^{\infty} f(t)dt + \int_T C_{\text{failure}}(t)f(t)dt \tag{14}$$

Consider an O&M cycle T_m. The expected duration of this O&M cycle $E(T_m)$ is

$$E(T_m) = \int_0^{\tau} tf(t)dt + \tau\int_{\tau}^{\infty} f(t)dt = \int_0^{\tau} tf(t)dt + \tau R(\tau) \tag{15}$$

The expected cost rate of maintenance $E[c_{om}(T_m)]$ for the m^{th} O&M cycle T_m, is therefore given by

$$E[c_{om}(T_m)] = \frac{E[C_{om}(T_m)]}{E(T_m)} \tag{16}$$

Power Plant Profit. The key elements that define the power plant profit are the value of power or gross revenue due to the sale of electricity and the cost of electricity, which includes cost of fuel, cost of operations and maintenance (excluding cost of fuel), and depreciation of the power plant. Gross revenue from selling electricity, GR, during time period T is given by

$$\text{GR} = \int_T M_p(t)[P(t)dt] \tag{17}$$

The cost of fuel $C_{\text{fuel}}(T)$ during time period T, and is given by

$$C_{\text{fuel}}(T) = \int_T F_c(t)P(t)\text{HR}(t)dt \tag{18}$$

The depreciation parameter accounts for the investment cost for the power plant. The depreciation of the power plant is given by:

$$Q(T) = \int_T q(t)dt \tag{19}$$

Let NR be the profit of a power plant. The expected profit of a power plant over the stated period of time T is therefore given by

$$\begin{aligned} E[\text{NR}(T)] = {} & \int_T [P(t)[M_p(t)] - F_c(t)[\text{HR}(t)]P(t)]dt \\ & - \left[C_{pm}\int_T^{\infty} f(t)dt + \int_T C_{\text{failure}}(t)[f(t)dt] + \int_T q(t)dt \right] \end{aligned} \tag{20}$$

The expected profit rate for the m^{th} O&M cycle T_m is therefore given by

$$E[\text{NR}(T_m)] = \frac{E[NR_{om}(T_m)]}{E(T_m)} \tag{21}$$

More details regarding O&M modeling are given in Ref. [16].

Problem Formulation. The planning horizon for preventive maintenance is an important issue, since the plant value is directly related to its consumed lifetime. The determination of the planning horizon therefore needs to take into account high-level plant owner strategies, such as power plant replacement strategy. A desired lifetime is defined here as eight O&M cycles. The problem is to determine the eight optimal preventive intervals $\{T_m\}, m = 1, 2, \ldots, 8$ for this power plant based on the projection of the long-term electric power market, power plant performance degradation, and operational risk. For such decision making, the trade-off between risk and reward, i.e., the significance of performance degradation, risk, and spark spread,[2] is very important.

One approach to evaluate the economic performance of the power plant is to calculate the profit rate or cost rate over its entire service life. In this situation, the profit rate or cost rate is calculated by summing up the cumulative profit or maintenance cost over its entire service life (eight O&M cycles), and dividing it by the entire service life of the power plant.

A second approach is to calculate the profit rate or cost rate of maintenance of each O&M cycle separately. This approach allows that the power plant economic performance be evaluated on a shorter term basis. This latter approach is employed in this study.

For preventive maintenance scheduling in this study, the optimization criteria employed includes both the expected profit rate, which is defined in Eq. (21), and the expected cost rate of maintenance, which is defined in Eq. (16), for each O&M cycle. The formulized profit-based sequential preventive maintenance scheduling problem is therefore given below:

Maximize:

$$E[\mathrm{NR}(T_m)] = \frac{E[NR_{om}(T_m)]}{E(T_m)}$$

By optimizing T_m, where $m=1,2,\ldots,8$.

The formulized cost-based sequential preventive maintenance scheduling problem is therefore given below:

Minimize:

$$E[c_{om}(T_m)] = \frac{E[C_{om}(T_m)]}{E(T_m)}$$

By optimizing T_m, where $m=1,2,\ldots,8$.

As discussed in Ref. [1], the profit-based approach relies on knowledge about the economic performance of the power plant. However, a projection of the future operating profile is necessary to evaluate power plant output and heat rate, performance degradation, and risk assessment. Also the projection of future electric power market parameters, such as price of electricity, is also necessary, since these factors are pertinent to the evaluation of power plant gross and net revenues.

The time horizon for this sequential preventive maintenance scheduling involves the entire service life of the gas turbine power plant, which is a time period up to even more than a decade. For the preventive maintenance scheduling problem involving such a long-term period, the variations of the electric power market in a relatively short-term timeline (i.e., daily variation) are essentially noise variables, and the detailed modeling of electric power market and weather conditions on daily basis are not necessary. However, models to predict the seasonal and long-term trends of the dynamic electric power market are important for effective profit-based preventive maintenance scheduling. Therefore, in this study, the daily variations of the electric power market and weather conditions are not taken into account. For simplicity, only the long-term trend of the price of electricity and price of fuel are modeled, with the seasonal trend of the electric power market and weather conditions not considered. For the same reason, a uniform future operating profile over the entire service life of the power plant is assumed.

The example shown in this paper is for illustration purposes only. The assumptions made for performance degradation and reliability of the power plant are rather conservative. Generally, the following assumptions for the sequential preventive maintenance problem are used:

1. The gas turbine power plant is brand new at the beginning of its service.

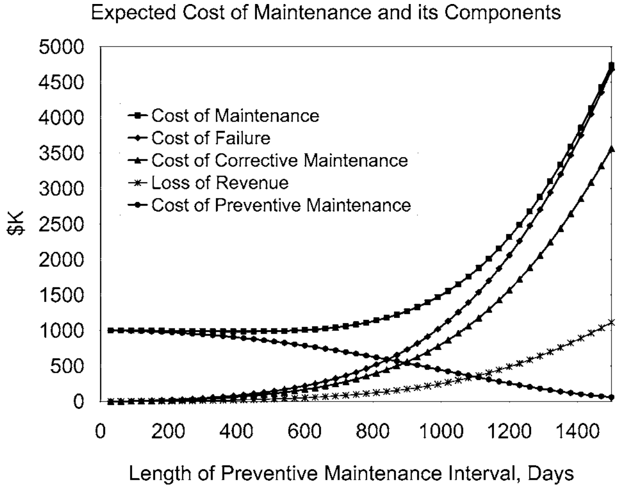


Fig. 1 Expected cost of maintenance and its components as a function of the length of the preventive maintenance interval for the first O&M cycle

2. The planning horizon is eight operations and maintenance cycles.
3. The preventive maintenance actions are performed at a sequence of intervals T_k, and preventive maintenance is imperfect maintenance, as defined previously.
4. Corrective maintenance is performed whenever the system fails, and the corrective maintenance is minimal maintenance, as defined previously.
5. The duration for preventive maintenance is one month.
6. The gas turbine system reliability functions, including hazard rate, probability density function, and reliability, are defined only when it is in operation
7. The gas turbine components and system reliability are Weibull distributions.

Numerical Results Analysis

Scenario Description. A numerical example is introduced here to demonstrate the feasibility of the proposed approach. In this example, both the profit-based and the cost-based sequential preventive maintenance approaches are employed to determine the eight optimal preventive intervals over the entire service life of a base load combined cycle power plant. A base load combined cycle power plant with a single gas turbine is investigated. This base load power plant runs 24 hr/day, continuously, during its normal operation. A uniform future operating profile over the entire service life of the power plant (which is base load, natural gas fuel, and without power augmentation) is assumed.

Preventive Maintenance Scheduling for First O&M cycle. A parametric study on the impact of the timing of the preventive maintenance schedule on power plant economic performance over each O&M cycle is performed; this is done by manipulating the maintenance interval for the O&M cycle. The power plant economic performance (which includes expected cost of maintenance, the expected cost of maintenance per unit time, the expected profit, and the expected profit per unit time) is investigated. The optimal timing of the preventive maintenance schedule is then determined, and the detailed operation for the power plant under this optimal preventive maintenance schedule is provided.

The variations of power plant expected cost of maintenance, revenue, and profit as functions of the length of the preventive maintenance interval are investigated. The length of the maintenance interval for the first O&M cycle varies from 30 days to 1500 days. The results are shown in Figs. 1–3.

The elements that are combined to determine the total cost of

---

[2] Here *spark spread* is defined as the difference between the spot market value of natural gas and the electricity at a given time based on the conversion efficiency of a given gas-fired plant.

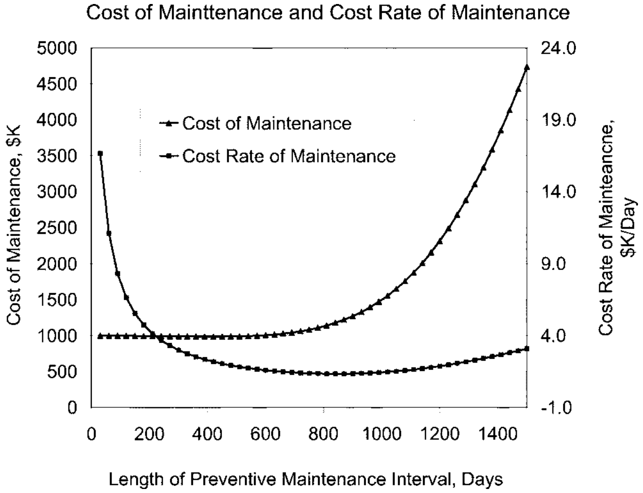


Fig. 2 Cost of maintenance and cost rate of maintenance as a function of the length of preventive maintenance interval for the first O&M cycle

maintenance are shown in Fig. 1 as a function of the preventive maintenance interval. Then in Fig. 2, it is shown that the expected cost rate of maintenance is very high when the maintenance interval is very small, say, 30 days. The cost of preventive maintenance dominates the total cost of maintenance, and the relatively short time span results in a high cost rate. As the maintenance interval increases, the expected cost of failure keeps increasing, and the expected cost of preventive maintenance keeps decreasing. As a result, the expected maintenance cost decreases slightly, reaches its minimum, and then climbs as the maintenance interval increases further. The cost rate of maintenance decreases rapidly as the maintenance interval increases, reaches it minimum, and then climbs. Please note that the optimal maintenance interval for the cost rate of maintenance lags behind that for the total cost of maintenance.

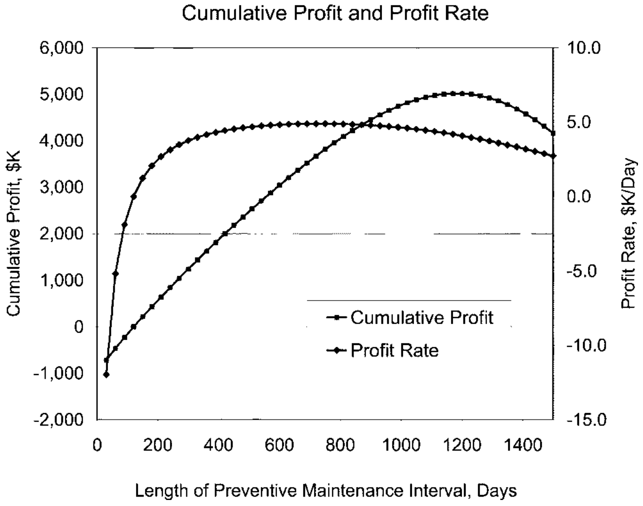


Fig. 3 Expected profit and profit rate as a function of the length of preventive maintenance interval for the first O&M cycle

Table 1 Results of profit-based sequential preventive maintenance scheduling

Preventive maintenance	Preventive maintenance intervals (days)	Date of scheduled preventive maintenance (day)
First	729	729th
Second	692	1451th
Third	622	2103th
Fourth	552	2685th
Fifth	490	3205th
Sixth	438	3673th
Seventh	393	4096th
Eighth	353	4479th

It is shown in Fig. 3 that the cumulative profit and, therefore, the profit rate (average profit per day) for the operations and maintenance cycle are negative, when the maintenance interval is very small, say, 30 days. The length of operation time period is so small that the revenue collected is less than the cost of operations and maintenance. As the maintenance interval increases, the expected revenue and cost of operations and maintenance keeps increasing, but the incremental revenue outweighs that of the cost of operation. As a result, the expected profit increases as the maintenance interval increases. It climbs up and reaches its maximum value. As the length of maintenance interval increases further, the incremental cost of operation outweighs the incremental revenue, and the value of cumulative profit goes down. The profit rate follows the same trend of the cumulative profit, however, the optimal maintenance interval for the profit rate is lower than that for the cumulative profit.

Sequential Preventive Maintenance Scheduling. The method to optimize the maintenance interval for a single O&M cycle can be used to sequentially optimize the preventive schedules over the entire service life of the power plant. A sequential preventive maintenance optimization, which determines the eight optimal maintenance intervals for the entire service life of the power plant, is then performed.

A series of eight preventive maintenance actions is scheduled using the profit rate for each O&M cycle as the objective function. The optimized preventive maintenance schedules are shown in Table 1, and the power plant aging, reliability and performance degradation, and cost and profit information under these optimal preventive maintenance schedules are shown in Figs. 4–8. The results show that as the power plant ages, the optimal maintenances interval becomes smaller. This is because as the unit becomes "older," the performance degradation and probability of failure of the power plant become more significant, i.e., the performance loss and the risk both increase.

The virtual age of the power plant is the result of partial restoration due to preventive maintenance, and the actual age and virtual age of the gas turbine power plant in factored fired hours are shown in Fig. 4. The failure rate of the gas turbine as a function of calendar time is shown in Fig. 5, and the performance degradation as a function of calendar time is shown in Fig. 6. These figures show that the gas turbine power plant is an aging system. The preventive maintenance is imperfect in that each preventive maintenance action partially reduces the age of the gas turbine. Therefore, the reliability and performance degradation are partially restored whenever a preventive maintenance is performed.

The cumulative cost of maintenance and its components, which include the cost of preventive maintenance, the loss of revenue due to unavailability of the plant, the cost of corrective maintenance to repair the plant, and the cost of failure of the damage due to failure, are shown in Fig. 7. It is shown that the cost of preventive maintenance and, therefore, the cost of maintenance jump whenever a scheduled preventive maintenance action is performed.

The cumulative cash flow of the power plant, which includes the cost of maintenance, revenue, cumulative spark spread, profit, and depreciation, are shown in Fig. 8. A cost-based sequential preventive maintenance scheduling is also performed, and the ex-

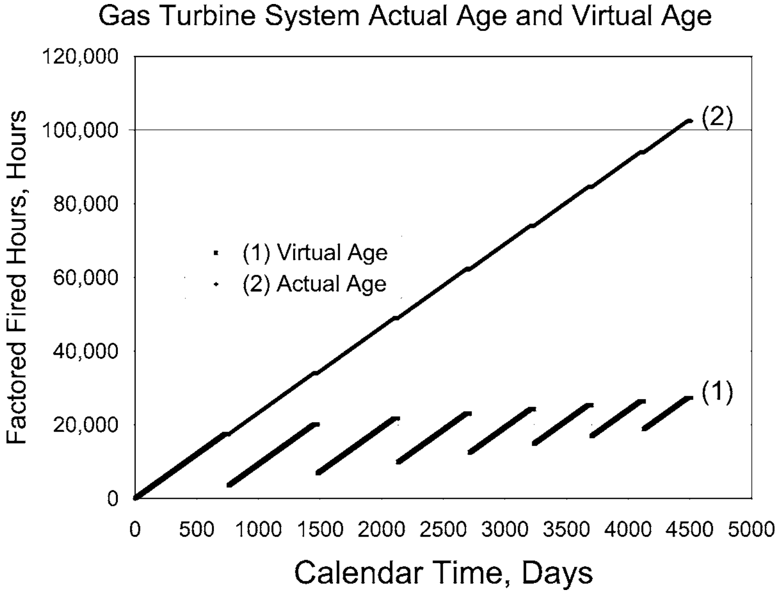


Fig. 4 Gas turbine system actual age and virtual age as a function of calendar time for the sequential preventive maintenance schedule

pected cost rates of maintenance for each O&M cycle is used as the objective function. The optimized preventive maintenance schedules are shown in Table 2.

Please note that the seasonal trends of the price of electricity, price of fuel, and weather conditions are not taken into account in the example introduced above, and a uniform future operating profile is assumed in this example. In actuality, however, the seasonal variations of the market signals are important factors, and the operating profile of the gas turbine power plant does change along the time line due to the dynamic electric power market.

To illustrate the impact of seasonal trends of the dynamic electric power market, the expected profit per day as a function of preventive maintenance interval with consideration of seasonal trend of price of electricity is created and is shown in Fig. 9. In this example, the price of electricity is assumed to be of a seasonal trend, and the price of electricity is higher in the summer than in the spring, fall, and winter because of high power demand in the summer. The results clearly show that the impact of the price of electricity on the profit rate of the power plant. The seasonal trend of price of electricity does affect the revenue profile and, therefore, the profit rate. This is different from the profit-rate distribution shown in Fig. 3. A more complicated pattern of profit-rate distribution is expected if the variation of future operating profile is taken into account. The mechanism of this effect is complicated. A full consideration of the dynamic electric power market, and, hence, the varying future operating profile, is therefore

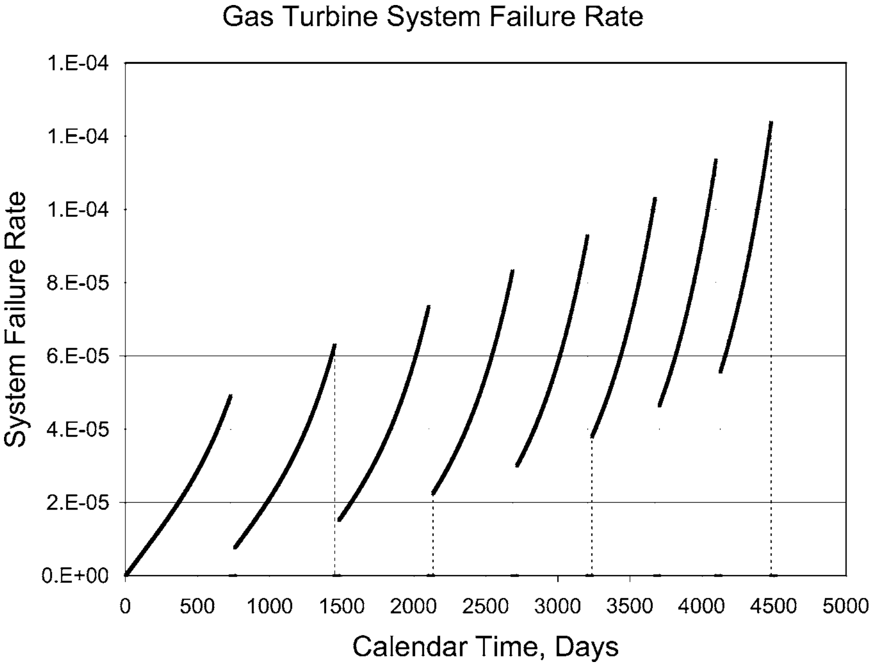


Fig. 5 Gas turbine system failure rate as a function of calendar time for the sequential preventive maintenance schedule

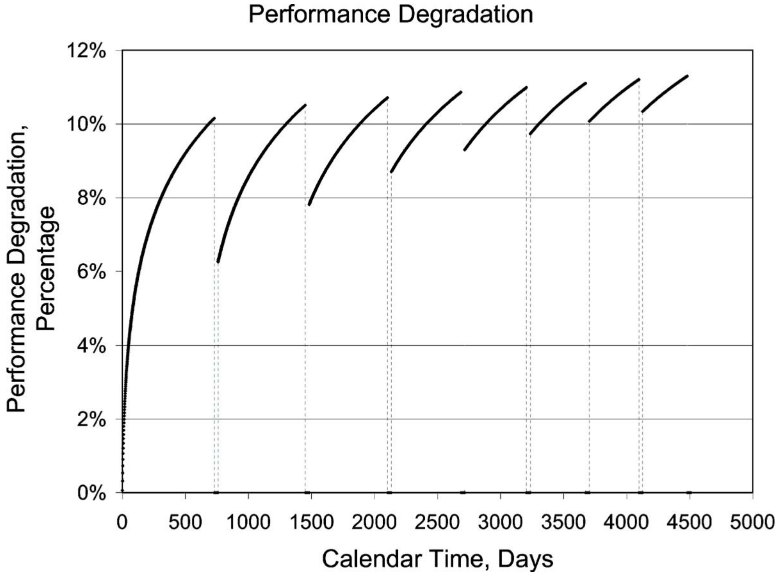


Fig. 6 Power plant performance degradation as a function of calendar time for the sequential preventive maintenance schedule

needed for more effective preventive maintenance scheduling. A profit-based methodology for gas turbine power plant outage planning has been developed to meet this need [1], and outage planning and long-term generation scheduling using this methodology is performed in a coupled fashion.

It must be mentioned that in the considered example the maintenance intervals are relatively shorter than those applied in current gas turbine power plant maintenance practices. This can be explained by the use of fairly conservative assumptions about the relevant maintenance parameters. In particular, the power plant performance and reliability degradation are assumed to be relatively fast paced, resulting in rapid performance loss and increased in operational risk.

Concluding Remarks

In this paper, a novel approach for maintenance scheduling of a gas-turbine-based power plant is introduced and a profit-based sequential preventive maintenance scheduling methodology is developed for more effective maintenance scheduling. A conceptual level study is performed, and a numerical example for profit-based sequential preventive maintenance scheduling is introduced. The procedure is implemented using a base load combined cycle power plant with a single gas turbine, and the results demonstrate the feasibility of the proposed approach. Sequential preventive maintenance planning is performed for the gas turbine power plant with eight operations and maintenance cycles over its entire

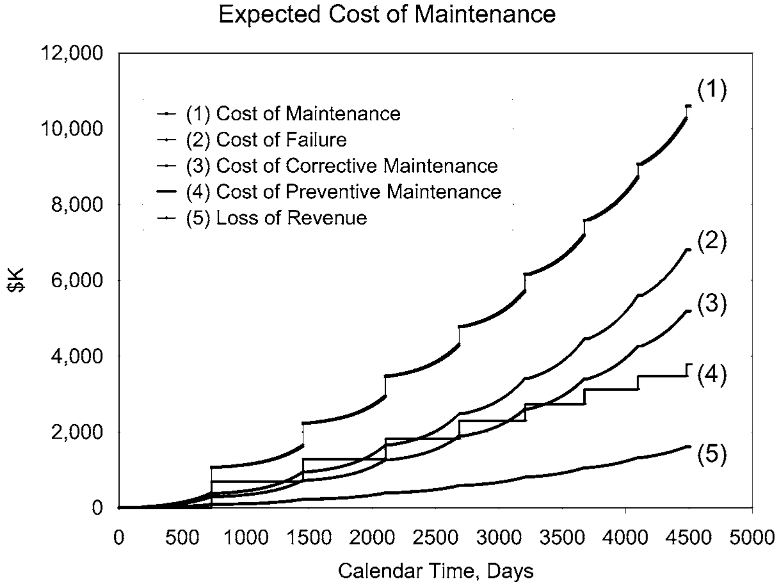


Fig. 7 Power plant expected cost of maintenance and its components as a function of calendar time for the sequential preventive maintenance schedule

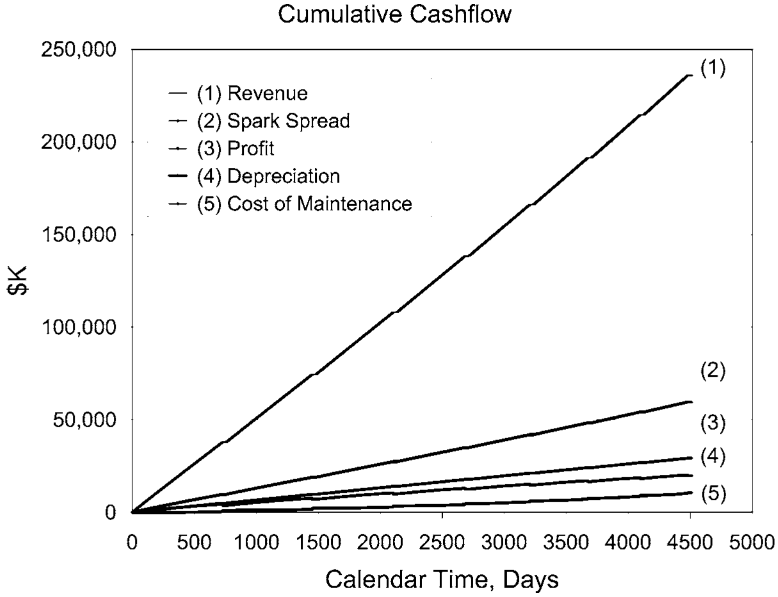


Fig. 8 Power plant cumulative cash flow as a function of calendar time for the sequential preventive maintenance schedule

Table 2 Results of cost-based sequential preventive maintenance scheduling

Preventive maintenance	Preventive maintenance intervals (days)	Date of scheduled preventive maintenance (day)
First	842	842th
Second	713	1585th
Third	609	2224th
Fourth	523	2777th
Fifth	455	3262th
Sixth	399	3691th
Seventh	351	4072th
Eighth	313	4415th

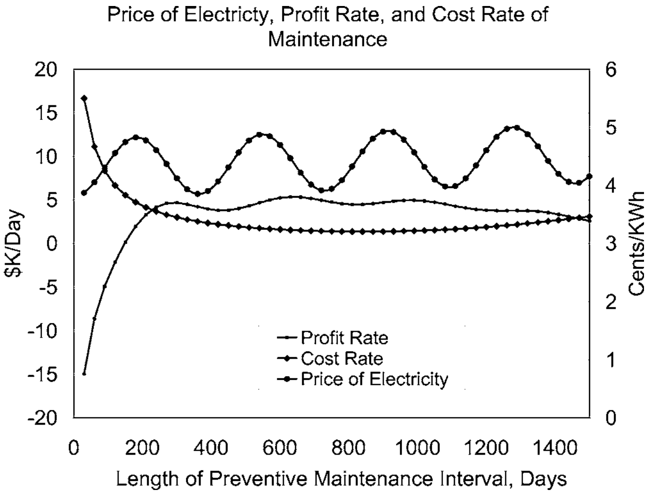


Fig. 9 Price of electricity, profit rate, and cost rate of maintenance as a function of the length of the preventive maintenance interval. (With consideration of the seasonal trend of the price of electricity.)

service life. The objective function for optimization is the profit rate or cost rate for each O&M cycle. The results show decreasing maintenance intervals as the power plant ages. By implication, new equipment should be more reliable with lower maintenance costs.

With the use of the profit-based sequential preventive maintenance scheduling, the power plant maintenance decisions depend not only on the maintenance cost, but also on the plant performance and the dynamic electric power market. Using this profit-based sequential approach instead of the traditional cost-based periodic preventive maintenance approach, it is expected that the cost of operations and maintenance will be reduced and the power plant profit will be increased.

It is understood that, in reality, many more factors are involved in the power plant maintenance scheduling, and the problem is much more complicated than the one addressed in this study. However, even this relatively simple example demonstrates the importance of effects that have not previously been modeled for gas turbine power plant sequential preventive maintenance scheduling. A more sophisticated method can be developed using the methodology presented in this paper for preventive maintenance scheduling for heavy-duty gas turbine power plants. With the implementation of these methods, improved profitability for gas turbine power plant systems is expected.

References

[1] Zhao, Y., Volovoi, V., Waters, M., and Mavris, D., 2006, "A Profit-Based Approach for Gas Turbine Power Plant Outage Planning," ASME J. Eng. Gas Turbines Power, **128**, pp. 806–814.
[2] Kelly, A., *Maintenance Strategy*, 1997 Reed Educational and Professional Publishing, Oxford.
[3] Hoeft, R., Janawitz, J., and Keck, R., "GER3620J-Heavy Duty Gas Turbine Operating and Maintenance Considerations," General Electric Energy.
[4] Lin, D., Zuo, M. J., and Yam, R. C. M., 2000, "General Sequential Imperfect Preventive Maintenance Models," Int. J. Reliab., Qual. Safety Eng., **7**(3), pp. 253–266.
[5] Wang, H., 2002, "A Survey of Maintenance Polices of Deteriorating Systems," Eur. J. Oper. Res. **139**, pp. 469–489.
[6] Nguyen, D. G., and Murthy, D. N. P., 1981, "Optimal Preventive Maintenance Polices for Repairable Systems," Oper. Res., **29**(6), pp. 1181–1194.
[7] Kobbacy, K. A. H., Proudlove, N. C., and Harper, M. A., 1995, "Towards an Intelligent Maintenance Optimization System," J. Oper. Res. Soc., **46**, pp. 831–853.
[8] Valdez-Flores, C., and Feldman, R. M., 1989, "A Survey of Preventive Maintenance Models for Stochastically Deteriorating Single-Unit Systems," Naval Res. Logistics Quart., **36**, pp. 419–446.

[9] Nakagawa, T., 1988, “Sequential Imperfect Preventive Maintenance Policies,” IEEE Trans. Device Mater. Reliab., **37**(3), pp. 295–298.
[10] Nakagawa, T., 1986, “Periodic and Sequential Preventive Maintenance Policies,” J. Appl. Probab., **23**, pp. 536–542.
[11] Liu, X., Makis, V., and Jardine, A. K. S., 1995, “A Replacement Model With Overhauls and Repairs,” Naval Res. Logistics Quart., **42**, pp. 1063–1079.
[12] Pham, H., and Wang, H., 1996, “Imperfect Repair,” Eur. J. Oper. Res., **94**, pp. 425–438.
[13] Barlow, R. E., and Hunter, L. C., 1960, “Optimal Preventive Maintenance Policies,” Oper. Res., **8**, pp. 90–100.
[14] Handlarski, J., 1980, “Mathematical Analysis of Preventive Maintenance Schemes,” J. Oper. Res. Soc., **31**, pp. 227–237.
[15] Yamin, H. Y., 2004, “Review on Method of Generation Scheduling in Electrical Power Systems,” Electr. Power Syst. Res., **69**, pp. 227–248.
[16] Zhao, Y., Volovoi, V., Waters, M., and Mavris, D., 2005, “An Integrated Framework for Power Plant Operational Modeling and Optimization,” ASME Paper No. GT-2005-69009.
[17] Diakunchak, I. S., 1991, “Performance Degradation in Industrial Gas Turbines,” ASME Paper No. 91-GT-228.
[18] Kijima, M., Morimura, H., and Suzuki, Y., 1988, “Periodical Replacement Problem Without Assuming Minimal Repair,” Eur. J. Oper. Res. **37**(2), pp. 194–203.

A Profit-Based Approach for Gas Turbine Power Plant Outage Planning

Yongjun Zhao

Vitali Volovoi

Mark Waters

Dimitri Mavris

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

Traditionally, for complicated systems, such as a gas turbine power plant, maintenance cost and online availability are two of the most important concerns to the equipment owner. However, in the deregulated electric power market, cost and reliability are not the only concerns. The ultimate goal is to maximize plant profitability, and this requires the evaluation of many different factors, including system performance, the aging and reliability of equipment, maintenance practices, and market dynamics accounting for the price and availability of fuel as well as the generation of revenues in competing markets. Thus, gas turbine power plant planning optimization is a complex problem, and comprehensive operational modeling and optimization methods are required. In this paper, a profit-based power plant outage planning approach is presented that reflects the new challenges posed by deregulation. Specifically, particular attention is paid to modeling power plant aging, performance degradation, reliability deterioration, and, importantly, the energy market dynamics. A multiple time-scale method is developed for coupled short-term, long-term generation scheduling and outage planning for this profit-based outage planning approach. The procedure is implemented for a base load combined cycle power plant with a single gas turbine, with the results demonstrating the feasibility of the proposed approach. [DOI: 10.1115/1.2179466]

Introduction

Performance requirements for modern heavy-duty gas turbines necessitate extreme operating conditions for hot gas path components. As a result, these critical components have a limited life span and, more generally, a gas turbine represents an *aging* system experiencing continuous degradation during its operation. This physical degradation manifests itself in performance degradation, as well as in an increased risk of forced outage. Operating conditions of gas turbines determine the aging processes (and degradation rates) of their components and therefore affect both reliability and performance degradation of the power plant. The most important factors influencing operating conditions include starting cycle, power setting, type of fuel, and level of steam or water injection [1].

Maintenance planning has a strong impact on profitability of a gas turbine power plant. Maintenance is the combination of all actions intended to maintain the plant or to restore it to a performance level so that it can perform its required functions [2]. Maintenance activities include inspection, repair, and replacement, and they constitute a significant proportion of the varying operating cost.

Timely scheduled (preventive) maintenance can offset the plant deterioration and partially restore/upgrade the system performance as well as improve its reliability by reducing the risk of component failures. On the other hand, preventive maintenance results in significant direct costs as well as in indirect costs due to the loss of revenue during the outage. A trade-off among these conflicting objectives comprises a problem of outage planning (i.e., determination of the timing of power plant shutdown for the next preventive maintenance) [3]. The problem is further complicated by the need to plan the outage in advance due to contractual constraints (to minimize loss of revenue) and logistical considerations (to conduct maintenance in a cost- and time-effective manner). Finally, seasonal variations in loss of revenue also contribute to the complexity of the problem. Ideally, preventive maintenance would be done in periods when the demand for electric power is low, typically, in the spring and fall months. Recent research on power plant maintenance optimization can be found in Refs. [4–8].

Historically, gas turbine maintenance has been based on a fixed time interval according to recommendations from the power plant supplier. Generally speaking, these recommendations tended to be fairly conservative as minimizing the failure risks carried both financial and reputationwise incentives for the supplier, whereas servicing frequent maintenance outages provided a substantial additional source of revenue. Deregulation dramatically changed the nature of the contractual service agreements, which effectively provided strong incentives for risk management (as described in the previous paragraph) rather than risk minimization. Since operating conditions for each gas turbine vary from site to site, and from unit to unit, a unit-specific maintenance approach is needed for effective gas turbine maintenance scheduling. For such an approach to be successful, accurate predictions of reliability and performance degradation for each gas turbine is necessary.

In addition, deregulation has also brought new and more complicated means for generating revenue that cannot be reduced to simple cost considerations. Revenue to a power producer can come from fixed contracts, which cover varying periods of time from months to years, or it may come from the spot market, which cover varying periods of time from days to weeks. Thus, revenue models—another feature of what may be called market dynamics—are a major part of the optimization problem. Although the importance of the market dynamics is well recognized in the problem of unit commitment [9], to date the issue has been largely ignored in outage planning.

The power plant maintenance planning problem is therefore a complex problem involving all of the issues mentioned above: system performance, reliability, operations, maintenance, environment, and market dynamics. The following interdisciplinary modules are pertinent to this profit based approach:

- power plant system performance and factors that affect

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the Journal of Engineering for Gas Turbines and Power. Manuscript received August 25, 2005; final manuscript received September 28, 2005. Review conducted by K. C. Hall. Paper presented at the ASME Turbo Expo 2005: Land, Sea, and Air, Reno, NV, June 6–9, 2005, Paper No. GT2005-69011.

this performance, including an ambient conditions model and a performance degradation model

- operation and scheduled maintenance considerations, including component and system reliability
- economic considerations, including power demand and supply, value of power, price of fuel, etc.

This paper implements a general procedure for integrated power plant modeling introduced in Ref. [10]. A dual time-scale method is employed to determine coupled optimal generation scheduling and outage planning for a single operations and maintenance cycle. A complementary study has been performed to evaluate gas turbine power plant maintenance scheduling in the context of multiple operations and maintenance cycles. Specifically, the impact of unit aging on maintenance frequency is investigated over the lifecycle of power plants. It is found that economic factors drive a maintenance schedule to more frequent repairs and increasing associated expenses as equipment ages [11].

Aging. In order to quantify aging based on the usage history, maintenance (service) factors are introduced. A baseline condition for operating hours is defined as a gas fuel unit operating at continuous duty and with no water or steam injection [1]. The maintenance factor for this baseline is defined as 1.0. For operation that differs from the baseline, maintenance factors reflect the appropriately adjusted level of required maintenance. In so doing, the influence of factors (such as fuel type and quality, firing temperature setting, and the amount of water or steam injection) are considered with regard to an hours-based criteria [1]. Similarly, maintenance factors for starts can also be defined.

Therefore, the maintenance factor converts the effects of operating conditions deviating from the baseline to that of the baseline. When a gas turbine unit is running in an operating state more severe than the baseline condition, the corresponding maintenance factors will be <1, and hence, the actual maintenance interval will be reduced.

The value of maintenance factors is obtained from engineering experience. Let $m_h(t)$ be the maintenance factor of operating hours at time t, and $m_s(i)$ the maintenance factor of start i. The equivalent life of a system can be defined using two types of matrices, one is the factored fired hours, and the other is factored starts. The factored fired hours H_f for operating period $T$is defined below

$$H_f = \int_T m_h(t)dt \tag{1}$$

For starts, assume there are N startups during the operating time period T. The factored starts are defined as

$$S_f = \sum_{i=1}^{N} m_s(i) \tag{2}$$

The age of the gas turbine L_{ISH} is therefore given by $L_{ISH} = (H_f, S_f)$.

Performance Degradation. Power plant performance (output rate and heat rate (HR)) is a function of power plant design, technology upgrades, operating mode, ambient conditions, and degree of degradation. Let $P(t)$ be the power output rate of the power plant and $\text{HR}(t)$ the heat rate. Power output rate and heat rate depend on the following parameters:

$$P(t) = P\begin{pmatrix}\text{system design, technology upgrade,}\\ 0_\text{p}(t),\ \text{T}_a(t),\ \text{degradation}\end{pmatrix} \tag{3}$$

$$\text{HR}(t) = \text{HR}\begin{pmatrix}\text{system design, technology upgrade,}\\ 0_\text{p}(t),\ \text{T}_a(t),\ \text{degradation}\end{pmatrix} \tag{4}$$

Here, O_P is operating profile type and $T_a(t)$ are ambient conditions.

For each particular power plant, the actual output rate and heat rate of the power plant must include the effects of the degradation. Let P_0 and HR_0 be the output rate and heat rate of the power plant at the beginning of its service life, when no performance degradation has occurred. Let $\Delta^P(t)$ be the degradation of output rate at time t and $\Delta^{\text{HR}}(t)$ be the degradation of heat rate at time t. The degraded output rate and heat rate can be calculated using the following equations:

$$P(t) = P_0[1 - \Delta^p(t)] \tag{5}$$

$$\text{HR}(t) = \text{HR}_0[1 - \Delta^{\text{HR}}(t)] \tag{6}$$

Performance degradation is a function of system design and unit usage history, with the latter including both unit operating history and maintenance activities. Here, performance degradation of the power plant is modeled as a function of its actual operating hours τ. Assume the engine is with age τ_0 when it enters its service. According to Ref. [12], the equation that defines the curve can be given in the form of

$$\Delta(\tau) = a\ln[b(\tau + \tau_0) + c] + \Delta_0 \tag{7}$$

where Δ is the percentage of performance loss, which includes both power output and heat rate. a, b, c and Δ_0 are parameters that depend on the configuration and usage history of the engine, and their values can be determined using historical data.

Risk. In this study, the gas turbine aging and reliability is defined only when the power plant is in operation. Assume the reliability of the investigated item is a three-parameter Weibull distribution, which is frequently used for reliability modeling in the industry [13]. The failure rate function is given by

$$h(\tau) = \frac{\beta(\tau - \tau_0)^{(\beta-1)}}{\eta^\beta} \tag{8}$$

where $\eta > 0$ is the scale parameter, $\beta > 0$ is the shape parameter, and τ_0 is the location parameter.

The risk of a component or system failure during a time period of operation T is defined below

$$\text{Risk}(T) \equiv \int_T C_{\text{failure}}(t) f(t)dt \tag{9}$$

where C_{failure} is the consequence of the failure and f the probability density function.

Let C_{pm} be the cost of preventive maintenance. The expected cost of preventive maintenance during time period T is

$$E[C_{pm}(T)] = C_{pm}\int_{\tau_{pm}}^{\infty} f(t)dt \tag{10}$$

Then the total expected profit for a period T can be calculated as

$$E(NR) = \int_T \{P(t)M_p(t) - F_c(t)\text{HR}(t)P(t)\}dt - \left\{C_{pm}\int_{\tau_m}^{\infty} f(\tau)d\tau + \int_0^{\tau_m} C_{\text{failure}}(\tau)\cdot f(\tau)d\tau + \int_0^{\tau_m} q(\tau)d\tau\right\} \tag{11}$$

Here, all terms are previously defined except that $\tau_m = (H_f, S_f)$ is accumulated age, $M_p(t)$ price of electricity, and $F_c(t)$ price of fuel.

The first term in the expected profit equation (11) is the integrated-over-time difference between the value of power and the cost of fuel, which can be calculated on a daily basis. This term is here referred to as cumulative spark spread. Spark spread is the difference between the spot market value of natural gas and the electricity at a given time based on the conversion efficiency of a given gas-fired plant. As the conversion efficiency becomes greater, the spread between the market value of the gas and that of power derived by burning the gas becomes wider. The second term, which has three parts inside the large brackets, is the expected cost of preventive maintenance, cost of failure, and the depreciation of the power plant in the operation period of time T. Each part in the second term in the profit equation can be calculated after the long-term generation scheduling is performed, which will determine the age of the system during the time period of operation.

Problem Formulation: A Profit-Based Multiple Time-Scale Outage Planning

Coupling of Long-Term Generation Scheduling and Outage Planning. The profit-based outage planning approach relies on the knowledge about economic performance of the power plant. However, a projection of the future operating profile is necessary to evaluate power plant output and heat rate, performance degradation and risk assessment, and these factors are pertinent to the evaluation of power plant gross and net revenues. For effective outage planning, a projection of the unit usage, which depends on future electric power market and weather conditions in a relatively long-term future time horizon, is required. A profit-based outage planning approach therefore requires long-term generation scheduling.

The simplest approach, which is used in current preventive maintenance planning procedures, is to assume that the operating profile over the time horizon of interest is uniform. This approach is easy to implement and therefore extensively used in current engineering practice. In actuality, however, in the market-based operating environment, the operating profile shows strong variation due to market dynamics. An inaccurate uniform operating profile assumption leads to inaccurate system degradation estimates and, therefore, an ineffective outage plan. Thus, a methodology that is capable of capturing the variation of a future operating profile on a long-term basis is necessary for effective outage planning.

Scenario Description. In the scenario considered for this study, a base load combined cycle power plant with single gas turbine is investigated. For this base load gas turbine based power plant, it is assumed that two major preventive maintenance procedures, i.e., a combustion inspection and a hot gas path inspection or major inspection, are scheduled in every three years of operation. It is therefore assumed in this study that the operations and maintenance cycle for this power plant is 18 months. It is also assumed, in the beginning of the time period of concern, that the gas turbine has an initial age of 5000 factored fired hours after the last major preventive maintenance. The next scheduled preventive maintenance, which is a hot gas path inspection, is scheduled in the eighth month and the duration of the maintenance is one month. Furthermore, it is projected that there is a peak demand (wide spark spread) during the month of scheduled maintenance, so it might be advisable to shift the prescheduled hot gas path inspection to some other time period in order to take advantage of the wide spark spread. For such decision making, the trade-off between risk and reward (i.e., the significance of performance degradation, risk, and spark spread), is very important. In this outage departure problem, the timing of the outage for the next preventive maintenance is selected in such a way that the overall expected profit of the power plant during an operations and maintenance cycle is maximized.

Long-Term Generation Scheduling Using a Dual Time-Scale Approach

General Method. One of the most challenging problems in the electric power generation business is balancing short-term productivity with the optimal level of production over a long time period. At the level of a single power plant, there are a significant number of control variables that affect the operation of a power plant and its profitability. Most of the involved variables require short-term (weekly or daily) assessment and the corresponding optimization problems are addressed at this small time scale, i.e., the operator strives to optimize the profits at any given point in time given constraints, demand, and pricing environment. A full-blown long-term optimization of an operating profile is not practical at the same level of detail due to the size of the problem [14–16]. Also, the detailed scheduling of long-term operation on a daily basis is not reasonable due to the limited accuracy of long-term energy market projection.

A dual time-scale method for solving the long-term generation scheduling problem is introduced in Ref. [17]. The dual time-scale approach allows combining the detailed granularity of the day-to-day operations with global (seasonal) trends, while keeping the resulting optimization model relatively compact. Furthermore, this dual time-scale approach can incorporate gas turbine performance, the dynamic electric power market, long-term power plant generation scheduling, and outage planning. A brief introduction to the dual time-scale long-term generation method is introduced here as follows.

The objective is to maximize the long-term profitability of a gas turbine power plant by optimizing the operating profile of gas turbine operation under a dynamic environment, in which the value of power, price of fuel, and plant operating condition are stochastic in nature. The optimization problem is solved in two steps: first, a local (e.g., a single-day) optimization problem is solved parametrically where accumulated equivalent starts and fired hours are fixed. These two parameters are considered to be the two major factors affecting scheduled maintenance. As a result, the objective (e.g., net revenue) at the local level is expressed as a function of equivalent starts and fired hours, while all the actual plant control variables are embedded (hidden) as a result of the local optimization. Next, a “global” (a time period of certain length until the next scheduled preventive maintenance) problem is posed, where there are only two unknowns (equivalent starts and fired hours) per local time segment.

Operating Profile. The operating profile is a control variable, and it is defined on a daily basis. A typical operating profile type defines the starts setting, load setting, fuel type, and power augmentation. The start setting has options, such as hot starts, cold starts, and emergency starts. The load setting determines if the system is operating in base load, peak load, or part load. The type of fuel can be natural gas, liquid fuel, etc. Power augmentation defines if steam or water injection is employed. Each combination of these parameters defines an operating profile. To reduce the scale of the problem, operating parameters are converted into a compact description of various scenarios for the daily operation profile of a gas turbine. An example of possible operating profiles for continuous operation is shown in Table 1. Maintenance factors are established for each operating profile.

It is understood that, in actual engineering practices, a more extensive investigation of the various operating parameters and their corresponding maintenance factors, which affect the life of various components of gas turbine power plants, have to be modeled for effective operational planning. To demonstrate the general method, however, only two operating parameters—load setting and power augmentation—are investigated in this study. Although in this simple case all possible combinations are considered (Table 2), in general, a design of experiments (DOE) methodology is employed to capture the dependence on operating parameters.

The maintenance factors for each operating profile type are also provided in Table 2. These maintenance factors are normalized for the purpose of illustration.

Modeling of Price of Electricity and Weather Conditions. The variation of ambient temperature for 12 types of day is shown in Fig. 1. The variation of ambient temperature includes the daily

Table 1 Parameters for continuous operation

Load setting	Fuel type	Steam injection	Water injection
Base load	Natural gas	On	Off
		Off	On
		Off	Off
	Distillate oil	On	Off
		Off	On
		Off	Off
	Heavy fuel	On	Off
		Off	On
		Off	Off
Peak load	Natural gas	On	Off
		Off	On
		Off	Off
	Distillate oil	On	Off
		Off	On
		Off	Off
	Heavy fuel	On	Off
		Off	On
		Off	Off

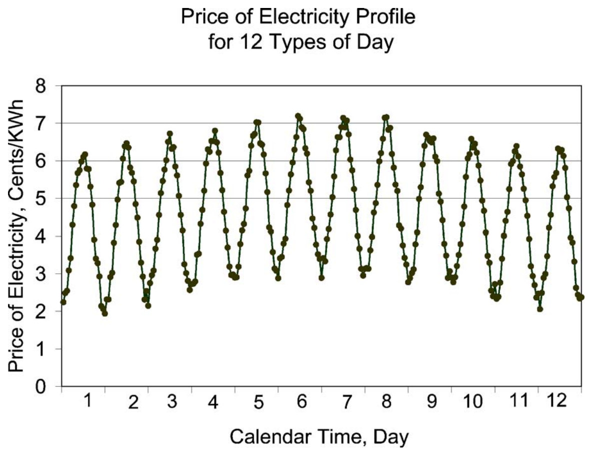


Fig. 2 Yearly variation of price of electricity

variation and seasonal variation. It is assumed the ambient temperature is relatively low in the early morning, keeps increasing until noon, then decreases and reaches the minimum at midnight. The seasonal variation shows that the average ambient temperature is relatively low in the spring, keeps increasing in the summer, and then decreases in the fall and reaches the minimum in the winter. Random factors are used to model the stochastic nature of ambient temperature.

Similarly, the daily variation of price of electricity is shown in Fig. 2. It is assumed that the price of electricity is lower between midnight and early morning than during the day, and the price of electricity is higher in the summer than in the spring, fall, and winter because of high power demand in the summer. Please note these assumptions do not necessarily match actuality, and what is important here is the variation in a time line.

Daily Time-Scale Optimization (Local Optimization). The strategy for local optimization consists of a separate optimization for each profile with respect to its own parameters followed by a selection of the best profile. The purpose of daily time-scale optimization is to construct optimal daily cumulative spark spread profiles as functions of daily usage of the power plant, i.e., daily factored fired hours and factored starts.

Table 2 Simplified operating profiles and maintenance factors

Operating profile types	Load mode	Steam injection	Maintenance factor
1	Base	Off	1
2	Base	On	1.5
3	Peak	Off	2
4	Peak	On	2.5

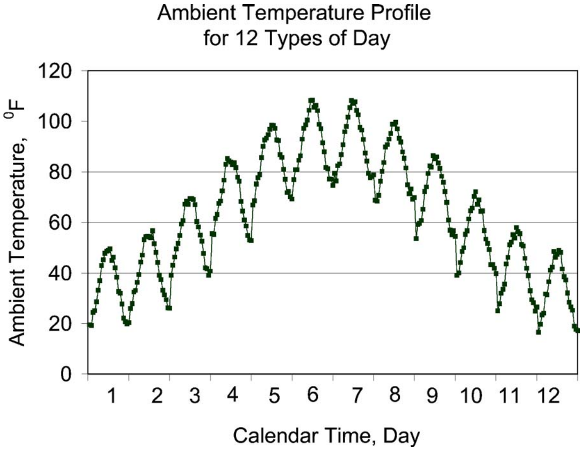


Fig. 1 Yearly variation of ambient temperature

Spark spread is determined by price of electricity, price of fuel, and power plant heat rate. Spark spread SS is calculated using the following equation:

$$\mathrm{SS}(d,t) = 10[M_p(d,t)] - F_c(d)\frac{HR(d,t)}{1000} \tag{12}$$

The units here for price of electricity, price of fuel, heat rate, and spark spread are cents per kilowatt-hour, dollars per million BTUs, BTUs per kilowatt-hour, and dollars per megawatt hour, respectively. The spark spread depends on the operating profile type O_P since it is a function of the heat rate of the power plant, which is a function of its operating profile type.

The spark spread as a function of calendar time for 12 types of day when the unit is running under operating profile type 1 (base load without power augmentation is shown in Fig. 3. As expected, the spark-spread profile follows the same trend as that of the price of electricity.

Day type 1 is a typical day in January in the winter, when it is assumed the demand of electric power is relatively low and therefore the average price of electricity is low (the situation might be vary depending on a geographical location [18]). The price of electricity in the early morning and midnight is so low that fuel cost is higher than the revenue produced from the sale of electricity. As a result, the spark spread is negative for that time period,

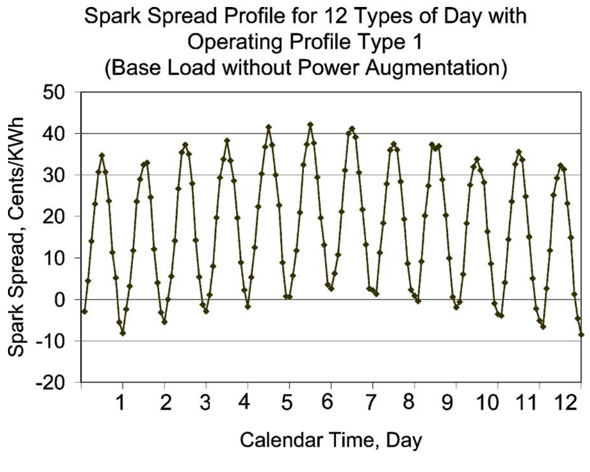


Fig. 3 Spark spread as a function of calendar time for the year

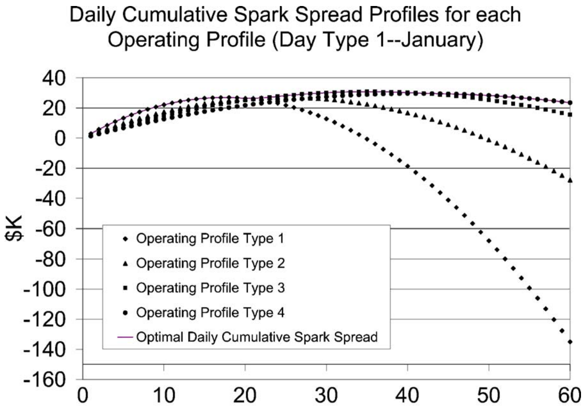


Fig. 4 Cumulative spark-spread profiles—day 1

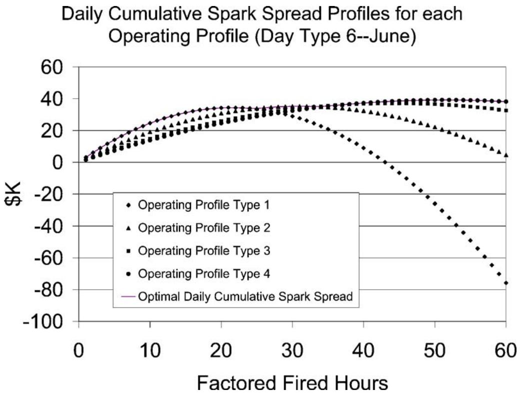


Fig. 5 Cumulative spark-spread profiles—day 6

which means money is lost if the power plant is turned on. As the season shifts from spring into summer, the demand for electric power increases and, hence, the price of electricity, and therefore the spark spread becomes wider. This is shown in day types 6 and 7, which are for a typical day in the summer. In this case, the spark spread in a summer day is always positive, even in the early morning and midnight. This means the power plant is making a profit as long as the plant is in operation.

A daily cumulative spark spread (DSS) is defined and calculated along the time line of daily operation. The daily cumulative spark spread is given by Eq. (13), which is the difference between the daily gross revenue of selling electricity and daily cost of fuel

$$\mathrm{DSS} = \int_0^{24} P(t)\left\{10[M_p(d,t)] - F_c(d)\frac{HR(d,t)}{1000}\right\}dt \qquad (13)$$

The parameter DSS depends on operating profile type and actual daily operating time because the heat rate is a function of operating profile type. But, in addition, it also depends on the extraneous parameters, including price of electricity $M_p(d,t)$, price of fuel $F_c(d)$, and ambient conditions $T_a(d,t)$. Note that for a given day during local optimization the ambient condition, price of electricity, and price of fuel are fixed based on forecasting data. The revenue and fuel cost of a plant depend only on the operating profile and actual fired hours H_a. Therefore, DSS is parametrically expressed as follows:

$$\mathrm{DSS} = \mathrm{DSS}[d,O_p,H_a(d)] \qquad (14)$$

Factored fired hours and factored starts are used as intermediate variables that link long-term generation planning and daily generation scheduling. The daily cumulative spark spread can therefore be expressed as a function of type of day, type of operating profile, and factored fired hours as given by

$$\mathrm{DSS} = \mathrm{DSS}[d,O_p,H_f(d)] \qquad (15)$$

The daily cumulative spark spread is integrated along the operating time line. As a result, a revenue profile for each operating profile type on a given type of day is calculated as a function of factored fired hours. The daily cumulative spark spread profiles for day type 1 and day type 6 are shown in Figs. 4 and 5, respectively.

The cumulative spark-spread profiles for day type 1 is shown in Fig. 4. Day type 1 is a typical day when the price of electricity is relatively low. For each operating profile, as the operating time increases (increases in factored fired hours), the cumulative spark spread increases due to positive spark spread and reaches the maximum value. It then decreases due to negative spark spread. In this case, the daily cumulative spark spread is not wide enough to justify running the power plant 24 h/day. Actually, as shown in Fig. 4, there is an optimal operating time for each operating profile that optimizes cumulative spark spread for each day, and less cumulative spark spread will be achieved if the plant is run for more time than that optimal operating time.

The data given in Figs. 4 and 5 show a trend that more cumulative spark spread can be achieved as the time of year shifts from the spring into the summer. As shown in Fig. 3, the spark spread in a summer day (day type 6) is always positive. For this reason, the power producers tend to run the plant more time in summer than in the spring and therefore produce more cumulative spark spread each day.

For each set of given day and daily factored fired hours H_f, an optimal operating profile can be identified that maximizes daily cumulative spark spread. The optimization is formulated as follows:

$$\mathrm{DSS}^* = \mathrm{DSS}^*[d,H_f(d)] = \max_{\text{Profile}}[d,O_p,H_f(d)] \qquad (16)$$

Here DSS^* is the constrained daily cumulative spark spread.

As a result, for each given day, the optimal daily cumulative spark-spread profile is constructed as a function of daily unit usage, i.e., the daily factored fired hours. Thus, an optimal daily cumulative spark-spread profile can be constructed for each day. The optimal daily cumulative spark-spread profiles as function of factored fired hours for day types 1–6 are shown in Fig. 6.

Yearly Time-Scale Optimization (Long-Term Generation Scheduling). The daily (short-term) optimization requirement is to maximize daily cumulative spark spread, while the yearly (long-term) optimization requirement is to maximize the cumulative

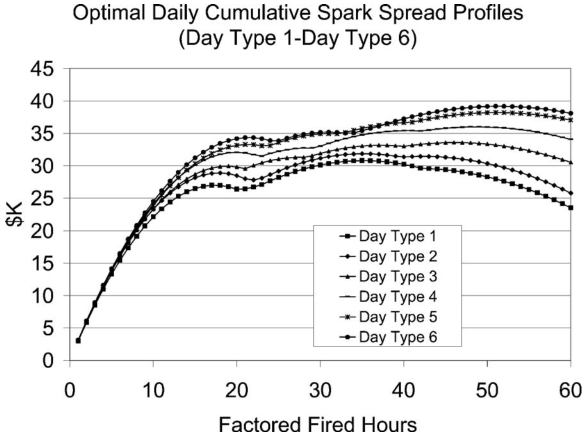


Fig. 6 Optimal daily cumulative spark-spread profiles (day type 1–day type 6)

profit of the power plant with consideration of long-term expected cost of maintenance and depreciation. This is done by optimizing the long-term generation scheduling for the given time period of operation.

The power plant will operate until an outage is scheduled for plant maintenance. The long-term economic performance of a plant is the integration of its daily performance over the long-term period. Assume there are a number of D_m days in the operation period T_m, i.e., the next outage is scheduled D_m days away from the current time. For a given future operation profile along the operation period T_m, the aging and, consequently, the degradation, risk, and depreciation of the power plant can be evaluated, and the expected cost of preventive maintenance, cost of failure, and depreciation can be determined.

For a particular future operating profile, suppose the accumulated age (factored fired hours and factored starts) over the operation period T_m is $\tau_m=(H_f,S_f)$, where H_f and S_f are defined by Eqs. (1) and (2). The expected cumulative spark spread CSS over the operation period T_m is

$$E(\mathrm{CSS})=\sum_{d=1}^{D_m}\mathrm{DSS}^{*}[d,H_f(d)] \tag{17}$$

Here, $\mathrm{DSS}[d,H_f(d),S_f(d)]$ is the optimized daily cumulative spark-spread profile for each day.

Considering Eqs. (9) and (10), the expected cost of failure is given by Eq. (18) and cost of preventive maintenance is given by Eq. (19)

$$\int_0^{\tau_m} C_{\mathrm{failure}}(\tau)[f(\tau)d\tau]=C_{\mathrm{failure}}\left(1-R\left(\sum_{d=1}^{D_m}H_f(d),\sum_{d=1}^{D_m}S_f(d)\right)\right) \tag{18}$$

Cost of preventive maintenance is

$$C_{pm}\int_{\tau_m}^{\infty} f(\tau)d\tau=C_{pm}R\left(\sum_{d=1}^{D_m}H_f(d),\sum_{d=1}^{D_m}S_f(d)\right) \tag{19}$$

Here $R(t)=1-\int_0^t f(\tau)d\tau$ is the plant system reliability. The depreciation function Q depends on the power plant design and configuration, and it is a function of the age τ of the power plant

$$Q(\tau)=Q\left(\sum_{d=1}^{D_m}H_f(d),\sum_{d=1}^{D_m}S_f(d)\right) \tag{20}$$

The expected profit equation is therefore given by

$$E(NR)=\left(\sum_{d=1}^{D_m}DSS^{*}[d,H_f(d)]\right)-\left[\begin{array}{l}C_{\mathrm{failure}}\left[1-R\left(\sum_{d=1}^{D_m}H_f(d),\sum_{d=1}^{D_m}S_f(d)\right)\right]\\ +C_{pm}R\left(\sum_{d=1}^{D_m}H_f(d),\sum_{d=1}^{D_m}S_f(d)\right)+Q\left(\sum_{d=1}^{D_m}H_f(d),\sum_{d=1}^{D_m}S_f(d)\right)\end{array}\right] \tag{21}$$

The problem now becomes how to assign factored fired hours and factored starts for each day in order to achieve optimized long-term payback for a given time period of operation. The global problem can be further reduced if individual days of operation with similar characteristics are grouped together. For example, Nixon et al. used 12 types of days to represent 12 months in a year, based on historic weather data [18]. In this paper, in order to reduce the number of variables, it is assumed there are $k=12$ segments of "representative" days at various time of each year. Each segment actually represents a month in a calendar year. Not only are operating conditions assumed to be similar on a representative day, but the global policies are similar as well. In such a setting, the 365 days of each year are mapped into the k types, with $n(d)$ days for each type, which is the number of days of each month. Now, at this point, our $2k$ optimization parameters are $H_f(d)$, $S_f(d)$, where $d=1,2,\ldots,k$.

The formulized yearly time-scale optimization follows. For a given time period of operation T_m with a number of days D_m, maximize

$$NR^{*}=\max_{H_f(d),S_f(d)}\left(\begin{array}{l}\left(\sum_{d=1}^{D_m}\mathrm{DSS}^{*}[d,H_f(d)]\right)\\ -\left[\begin{array}{l}C_{\mathrm{failure}}\left[1-R\left(\sum_{d=1}^{D_m}H_f(d),\sum_{d=1}^{D_m}S_f(d)\right)\right]\\ +C_{pm}R\left(\sum_{d=1}^{D_m}H_f(d),\sum_{d=1}^{D_m}S_f(d)\right)+Q\left(\sum_{d=1}^{D_m}H_f(d),\sum_{d=1}^{D_m}S_f(d)\right)\end{array}\right]\end{array}\right) \tag{22}$$

subject to

$$0\leqslant H_f(d)\leqslant H_{f,\mathrm{daily}}^{\max}$$

$$0\leqslant S_f(d)\quad d=1,2,\ldots,k \tag{23}$$

where $H_{f,\mathrm{daily}}^{\max}$ is the maximum daily usage of factored fired hours for a given power plant. This corresponds to the cumulative factored fired hours per day when the power plant is operating in 24 hours per day in the operating profile, and it will result in the highest maintenance factor. $S_f(d)$ is a non-negative integer representing factored fired starts on day d.

Results for Long-Term Generation Scheduling. As an example, the next outage for preventive maintenance is scheduled eight months away from the current time of consideration (i.e., the next preventive maintenance is scheduled in August). It is assumed that not only are operating conditions assumed to be similar on each day in a month, but the global policies are similar as well on each day. As a result, there are 12 different types of days for the entire year. It is also assumed that the current time is in the beginning of the year. A long-term generation scheduling using the dual time-scale method is performed, and the optimized future operation profile for the coming 12 months is generated.

The optimized operating profile, the daily factored fired hours,

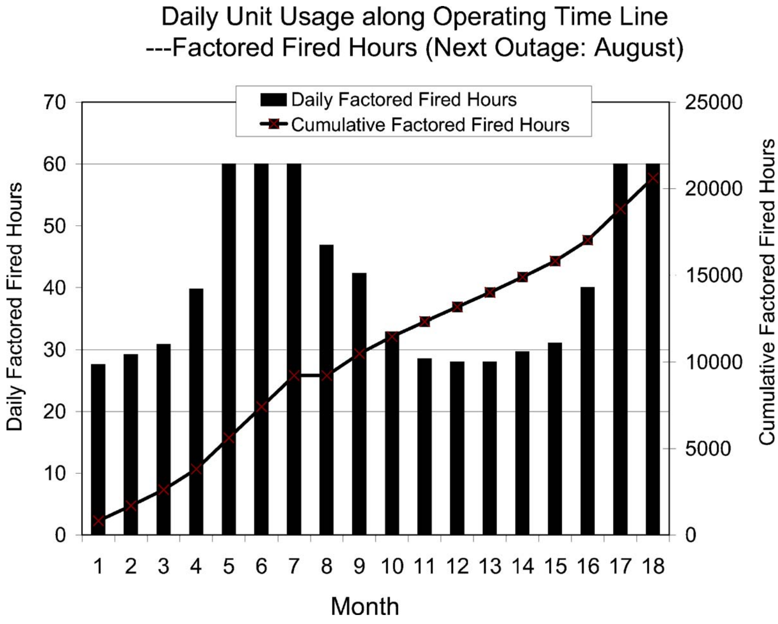


Fig. 7 Daily factored fired hours along the operating time line (next outage = eighth month)

and the daily actual fired hours for this operation and maintenance (O&M) cycle (from the first month to the eighteenth month) are shown in Figs. 7 and 8, respectively. The startup and shutdown schedule for each month is shown in Fig. 9.

It is found that the scheduled daily factored fired hours and daily actual fired hours follow the same trend as that of spark spread in the year; they increase as they go from spring into summer and decrease from fall into winter. As a result of the dynamics of the electric power market, the gas turbine is turned on to the highest output level during the summer and is scheduled to be in operation 24 h/day (i.e., the gas turbine is operating under peak load with steam injection (operating profile type 4)). During these months, the power plant is running continuously without shutdown. This is the case in the 5th, 6th, 7th, 17th, and 18th month. The power plant is scheduled to operate in a relative low output level, which is peak load without power augmentation (operating profile type 3), during the spring and the winter, and the power plant is started up and shut down on daily basis because the spark spread is not wide enough to justify 24 h of operation each day.

Profit Based Outage Planning. The dual time-scale long-term generation scheduling problem is a subproblem for the profit-based lifecycle-oriented outage planning problem. As a result of the optimization for long-term generation scheduling, the optimized expected profit equation is a function of the length of the time period of operation T_m (i.e., the number of days D_m)

$$NR^* = NR^*(D_m) \tag{24}$$

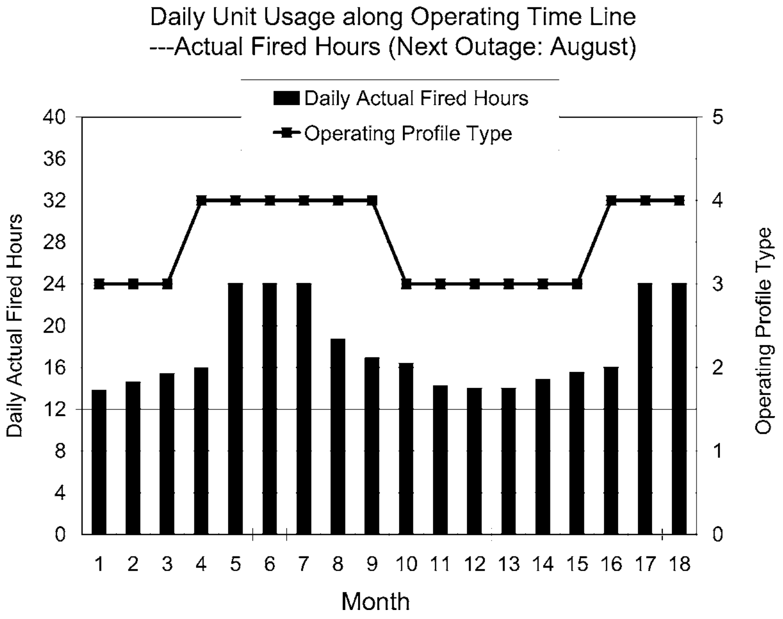


Fig. 8 Daily actual fired hours along the operating time line (next outage = eighth month)

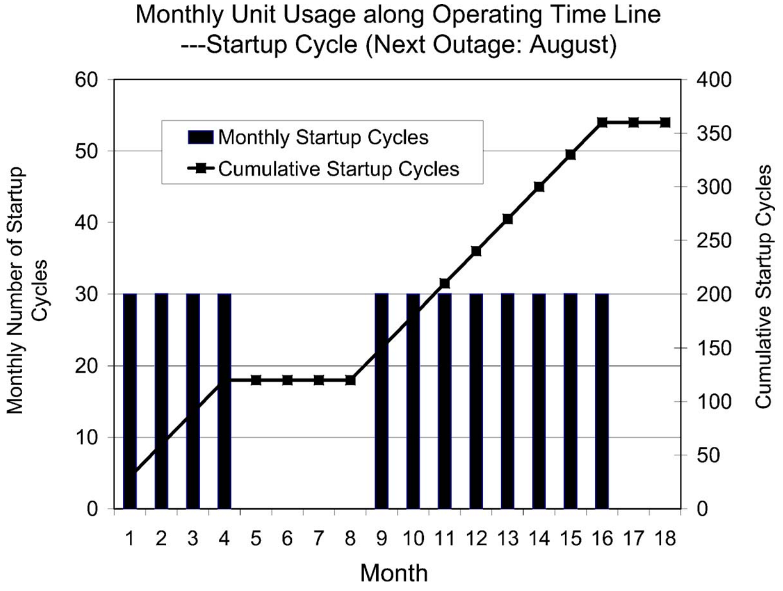


Fig. 9 Monthly unit startup and shutdown schedule along operating time line (next outage = eighth month)

The next optimization task is to maximize the expected power plant profit by optimizing the length of the operation period T_m (i.e., the number of days D_m). The outage optimization problem is therefore formulated below

Maximize:

$$NR^* = NR^*(D_m) \tag{25}$$

by optimizing the length of the operation period T_m, i.e., the number of days D_m.

The profit-based outage planning optimization is performed, and the results follow. The normalized expected profit in an O&M cycle as a function of outage schedule is shown in Fig. 10. It is found that, as the next preventive maintenance is postponed, the expected profit increases, and then it reaches the optimal. After that optimal point, the expected profit keeps decreasing if the preventive maintenance is postponed further. The results show that maximized profit over the 18 month O&M cycle is achieved when the next outage for preventive maintenance is performed in the ninth month, which is September in this case.

The market dynamics, performance, and reliability all interact simultaneously. The spark spread would drive the outage away from a given season when it is wide, when the power plant can gain much profit instantly, to a season with the smaller spark spread. This suggests that the next outage will occur most likely during the winter, particularly in the twelfth and thirteenth months. However, the impact of gas turbine aging is opposite. When the system is “young” enough, the performance degradation and the risk of forced outage is relatively less significant; thus, the incremental expected spark spread outweighs the incremental cost (the incremental risk and performance degradation). However, as the gas turbine system ages, the risk increases outpace the incremental profit. In this example, the performance degradation and risk associated with postponing the preventive maintenance from the ninth month (September) to the twelfth month (December) outweighs the marginal profit.

The detailed optimized generation schedule, including daily factored fired hours, operating profile type, and actual fired hours, for this outage schedule is shown in Table 3. This example clearly demonstrates that the optimal timing of power plant outage for preventive maintenance is influenced by the power plant perfor-

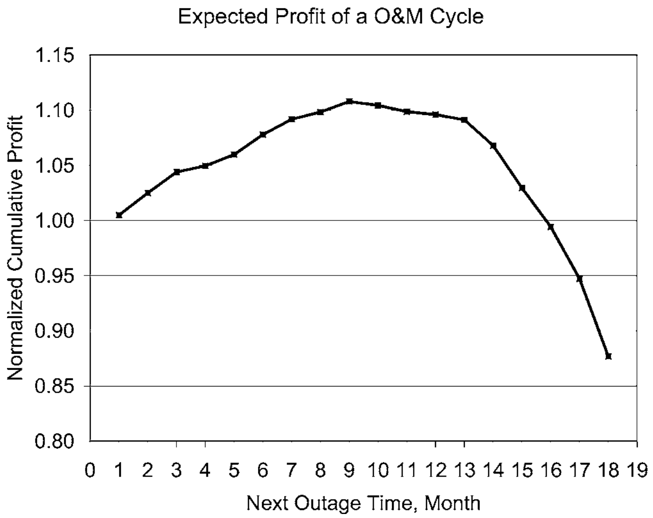


Fig. 10 Expected profit as a function of the next outage time

Table 3 The optimized generation schedule (the next outage is scheduled in the 9th month)

Month	Daily factored fired hours	Operating profile type	Daily actual fired hours
1	30.5	3	15.3
2	32.4	3	16.2
3	34.0	3	17.0
4	36.4	3	18.2
5	46.8	4	18.7
6	60.0	4	24.0
7	60.0	4	24.0
8	42.8	4	17.1
9	0.0	Outage	0.0
10	33.4	3	16.7
11	33.3	3	16.7
12	29.9	3	14.9
13	28.0	3	14.0
14	30.6	3	15.3
15	33.6	3	16.8
16	34.3	3	17.2
17	60.0	4	24.0
18	60.0	4	24.0

mance degradation, reliability, and market dynamics. As a result, outage planning that considers only performance and/or reliability will lead to a suboptimal solution. This provides a strong motivation for pursuing the profit based approach, where the performance, reliability, and market signals are considered in an integrated fashion.

Concluding Remarks

There is a need for profit-based outage planning for gas turbine power plants as a result of the deregulation of the electric power market. In this paper, a systematic approach for profit-based outage planning is introduced. The key factors for this profit-based approach include power plant aging, performance degradation, reliability deterioration, and, importantly, the energy market dynamics. Outage planning that considers only performance and/or reliability will essentially lead to a suboptimal solution.

A multiple time-scale operational scheduling method is developed for coupled generation scheduling and outage planning. The models that are currently being developed for this planning approach have been demonstrated in this paper in an example that uses a relatively simple power plant model operating over an 18 month period. It is found that this profit-based outage planning approach is capable of coupling power plant performance, reliability, and energy market dynamics, and therefore allows more effective outage planning with coupled generation scheduling. Using this multiple time-scale profit-based outage planning approach, an increase in the profitability of a gas turbine power plant is expected.

Several important enhancements of the developed procedure can be identified: a fleet-wide outage planning that takes into account coupled behavior of power plants due to both performance (e.g., fleet-wide performance requirements) and maintenance (e.g., constraints due to shared resources). To this end, a recently developed system modeling tool based on stochastic Petri Nets, SPN@ (stochastic Petri nets with aging tokens) is expected to provide a means to model relevant stochastic interactions in a compact and efficient manner [19]. Also, a more extensive modeling of the operating profiles and their corresponding maintenance factors would be helpful. Finally, more sophisticated cost models can further improve the quality of the decision making by incorporating other factors, such as power demand, power factor, taxes, insurance, etc. The price of electricity in the deregulated electric power market is set by rate structures as well as by the spot market. As a result, a high-fidelity modeling of time-varying price of electricity is a particularly challenging problem.

The method introduced in this paper along with the implementation of the above-mentioned enhancements represent a foundation for a comprehensive suite of software tools for facilitating effective outage planning in real-world situations and improving gas turbine power plant profitability.

References

[1] Hoeft, R., Janawitz, J., and Keck, R., 2003, "GER3620J-Heavy Duty Gas Turbine Operating and Maintenance Considerations," General Electric Energy.

[2] Kelly, A., 1997, *Maintenance Strategy*, Reed Educational and Professional Publishing, Oxford, Boston.

[3] Smith, D. J., 2003, "Unit Availability not Affected by Extending Outage Cycles," Power Eng. J., **107**(3), pp. 38–42.

[4] Tarbet, M., Takacs, L., and Sipos, S., 1996, "Optimizing the Interval for Turbine Generator Inspection Outages," ASME Joint Power Generation Conferences, ASME PWR-Vol. 30.

[5] Dewey, R., Roemer, M., Pollard, M., and McCloskey, T., 1998, "Optimization of Outage Interval for a Large Steam Turbine Unit," ASME International Joint Power Generation Conference, ASME PWR-Vol. 33.

[6] Latcovich, J., and Tanner, M., 1998, "Steam Turbine Risk Assessment Program (STRAP): Managing Steam Turbine Risk in Critical Applications," ASME International Joint Power Generation Conference, ASME PWR-Vol. 33.

[7] Endrenyi, J., Aboresheid, S., et al., 2001, "The Present Status of Maintenance Strategies and the Impact of Maintenance on Reliability," IEEE Trans. Power Syst., **16**(4).

[8] Sllva, P., Campanari, S., and Macchi, E., 2003, "Optimization of Operating Conditions and Compressor Cleaning Time Intervals of Combined Cycles in a Liberalized Market," ASME Paper No. GT2003-38455.

[9] Shahidehpour, M., Yamin, H., and Li, Z., 2002, *Market Operations in Electric Power Systems: Forecasting, Scheduling, and Risk Management*, Wiley, New York.

[10] Zhao, Y., Volovoi, V., Waters, M., and Mavris, D., 2005, "An Integrated Framework for Power Plant Operational Modeling and Optimization," ASME Paper No. GT 2005-69009.

[11] Zhao, Y., Volovoi, V., Waters, M., and Mavris, D., 2006, "A Sequential Approach for Gas Turbine Power Plant Preventive Maintenance Scheduling," ASME J. Eng. Gas Turbines Power, **128**, pp. 796–805.

[12] DiakunchakI. S., 1991, "Performance Degradation in Industrial Gas Turbines," ASME Paper No. 91-GT-228.

[13] Gertsbakh, I., 2000, *Reliability Theory With Applications to Preventive Maintenance*, Springer, New York.

[14] Hussain, K., 1991, "Solution Method for Unit Commitment Limitations and Utility Constraints," IEEE Comput. Appl. Power, **4**(1), pp. 16–20.

[15] Guan, X., Zhai, Q., and Papalexopoulos, A., 2003, "Optimization Based Methods for Unit Commitment: Lagrangian Relaxation Versus General Mixed Integer Programming," *IEEE Power Engineering Society General Meeting*, July 13–17, IEEE, New York, Vol. 2.

[16] Frauendorfer, K., Glavitsch, H., and Bacher, R., eds., 1993, *Optimization in Planning and Operation of Electric Power Systems*, Springer-Verlag, Berlin.

[17] Zhao, Y., Volovoi, V., Waters, M., and Mavris, D., 2004, "Power Plant Systems Operational Scheduling Using A Dual-Time Scale," 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, Albany, New York.

[18] Nixon, J. N., Waters, M., and Mavris, D., 2003, "An Evaluation of the Effect of Ambient Conditions on the Integration of Inlet Conditioning Systems with Industrial Gas Turbine Engines," *ASME Turbo Expo Proceedings*, ASME, New York, Vol. 2, pp. 821–832.

[19] Volovoi, V., 2006, "Stochastic Petri Nets Modeling Using SPN@," Proceedings of RAMS-2006 Symposium, Newport Beach, CA, January 26–29, Paper No. 2006RM-166.

Mustapha Chaker
Gas Turbine Division,
Mee Industries, Inc.,
204 West Pomona Avenue
Monrovia, CA 91016

Cyrus B. Meher-Homji
Principal Engineer,
Turbomachinery Group,
Bechtel Corporation,
3000 Post Oak Blvd., MS 73,
Houston, TX 77056-6503

Inlet Fogging of Gas Turbine Engines: Climatic Analysis of Gas Turbine Evaporative Cooling Potential of International Locations

Inlet fogging of gas turbine engines has attained considerable popularity due to the ease of installation and the relatively low first cost compared to other inlet cooling methods. With increasing demand for power and with shortages envisioned especially during the peak load times during the summers, there is a need to boost gas turbine power. There is a sizable evaporative cooling potential throughout the world when the climatic data is evaluated based on an analysis of coincident wet bulb and dry bulb information. These data are not readily available to plant users. In this paper, a detailed climatic analysis is made of 106 major locations over the world to provide the hours of cooling that can be obtained by direct evaporative cooling. This data will allow gas turbine operators to easily make an assessment of the economics of evaporative fogging. The paper also covers an introduction to direct evaporative cooling and the methodology and data analysis used to derive the cooling potential. Simulation runs have been made for gas turbine simple cycles showing effects of fogging for a GE Frame 7EA and a GE Frame 9FA Gas turbine for 60 and 50 Hz applications. [DOI: 10.1115/1.1707034]

1 Introduction

Gas turbine output is a strong function of the ambient air temperature with power output dropping by 0.54–0.9% for every 1°C rise in ambient temperature (0.3–0.5% per 1°F). On several heavy frame gas turbines, power output drops of around 20% can be experienced when ambient temperatures reach 35°C (95°F), coupled with a heat rate increase of about 5%. Aeroderivative gas turbines exhibit even a greater sensitivity to ambient conditions. A representation of the power boost capability for given inlet cooling potential for different types of gas turbines is shown in Fig. 1. This was derived using GTPRO[1] software over a range of turbines.

This loss in output presents a significant problem to utilities, cogenerators and independent power producers when electric demands are high during hot summer months. In the petrochemical and process industries, the reduction in output of mechanical drive gas turbines often curtails plant output. For example, at some liquefied natural gas (LNG) plants, production may have to be curtailed during the hot afternoons when the refrigeration capacity is limited by gas turbine driver power. One way to counter this drop is to cool the inlet air. While there are several cooling technologies available, fogging has seen large-scale application because of the advantage of low first cost when compared to other techniques including media evaporative cooling and refrigeration technologies.[2] Chaker et al. [1] has provided a detailed analysis of the evaporative cooling potential in terms of equivalent degree cooling hours (ECDH) for a large number of sites in the U.S.A. In this paper the same methodology is extended for 106 international locations outside the U.S.

A major obstacle faced by gas turbine users in analyzing the potential for fog evaporative cooling is that there is sparse climatic data available *in a form that users can make a decision on the benefits of evaporative cooling*. The obstacle may be broken into two factors:

1. Operators cannot easily locate the appropriate weather data for their site. Much of the data is available at a plant site may be based on *average* data points with no representation of the values of *coincident* dry and wet bulb temperatures. This data is invaluable when evaluating any evaporative cooling solution.
2. Even when some appropriate data is available through web sites or other sources, the data tables and information are not in a format to enables an operator to rapidly access the potential of evaporative cooling. The data has to be considerably massaged and collated before a meaningful estimate can be made of cooling potential at the site.

The object of this paper is to provide an easy to use solution to this problem providing a detailed analysis of multiple locations all over the world to allow users to evaluate power augmentation potential. This paper is intended to allow a rapid evaluation of site-specific cooling potential. McNeilly [2] has provided an excellent study on the importance of accurate climatic data when evaluating gas turbine inlet cooling projects. The relative potential of different gas turbines to capacity increase due to inlet cooling has been evaluated by Kitchen et al. [3].

In order to keep this paper self contained, some of the basic concepts of fog evaporative cooling are presented below.

---

[1] Program by Thermoflow, Inc.

[2] Cost ratios are about 5:1 but can vary based on project specifics.

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received December 2001, final manuscript received March 2002. Assoc. Editor: E. Benvenuti. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Amsterdam, The Netherlands, June 3–6, 2002; Paper No. 2002-GT-30559.

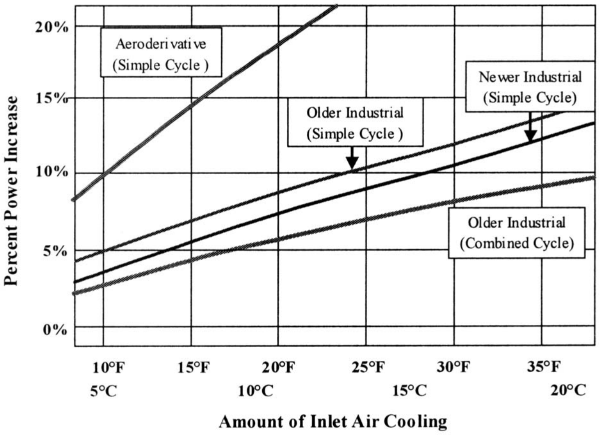


Fig. 1 Representation of power boost possible by inlet cooling

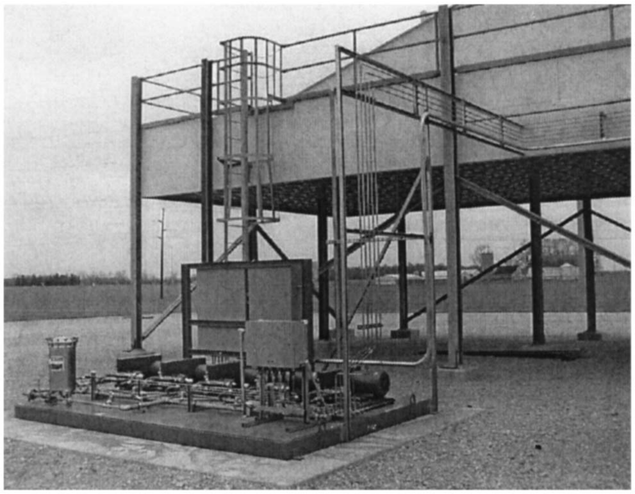


Fig. 2 Typical high pressure fogging skid. The feed lines from the high pressure pumps to the inlet system can be seen here.

2 Overview of Power Augmentation Strategies

There are several power augmentation strategies for gas turbines. These include:

- gas turbine inlet refrigeration-utilizing absorption or mechanical refrigeration,
- inlet fogging,
- HRSG supplemental firing—applies to combined cycle power plants (CCPP) only, and
- gas turbine water/steam injection.

A detailed study conducted by Tawney et al. [4] evaluated several options for power augmentation for combined cycle power plants. The results indicated that the option with the minimal EPC cost impact was inlet fogging. Inlet fogging was the only option that provided a small augmentation in heat rate, while the other options all worsened heat rate. In terms of return on equity, the highest return on investment was obtained by the combination of inlet fogging and supplemental firing of the HRSG. As a practical matter, several CCPPs are adopting fogging as a power augmentation strategy. This trend is being noted not only in the U.S.A. but in several parts of the world. Jones and Jacobs [5] have also studied various power enhancement techniques of combined cycle power plants.

3 Overview of Evaporative Cooling Technologies

3.1 Traditional Media Based Evaporative Cooling Technology. Traditional media based evaporative coolers have been widely used in the gas turbine industry especially in hot arid areas. The basic principle of evaporative cooling is that as water evaporates, it cools the air because of the latent heat of vaporization.

Traditional evaporative coolers are described in detail by Johnson, [6].

Evaporative cooler effectiveness is given by

$$E=\frac{T_{1\mathrm{DB}}-T_{2\mathrm{DB}}}{T_{1\mathrm{DB}}-T_{2\mathrm{WB}}} \tag{1}$$

where

T_1 = inlet temperature
T_2 = exit temperature of evaporative cooler
DB = dry bulb
WB = wet bulb.

A typical value for effectiveness is 85–90%, which implies that the wet bulb temperature can never be attained.
The temperature drop assuming an effectiveness of 0.9, is given by

$$\Delta T_{\mathrm{DB}}=0.9(T_{1\mathrm{DB}}-T_{2\mathrm{WB}}). \tag{2}$$

A psychometric chart can be used to obtain the value of the WBT. The exact power increase depends on the particular machine type, site altitude, and ambient conditions.

The presence of a media type evaporative cooler inherently creates a pressure drop that results in a drop in turbine output. As a rough rule of thumb, a 1 in. WG (2.54 cm WG) increase in inlet duct losses will result in a 0.35–0.48% drop in power and a 0.12% increase in heat rate. These numbers would be somewhat higher for an aeroderivative machine. The key issue with a traditional media evaporative cooler is that this increased pressure drop loss occurs year round even when the evaporative cooler is not in use. Increases in inlet duct differential pressure will cause a reduction of compressor mass flow and engine operating pressure. Increase in inlet differential pressure results in a reduction of the turbine expansion ratio.

The inherent loss of efficiency and increased inlet pressure loss in a traditional evaporative cooling system never allows for the maximum cooling effect to be attained. Water quality requirements are, however, less stringent than those required for direct fog cooling systems.

3.2 Inlet Fogging Technology. Direct inlet fogging is a method of cooling where demineralized water is converted into a fog by means of special atomizing nozzles operating at 138 bar (2000 psi). Details pertaining to the thermodynamics and practical aspects of fogging have been described in Meher-Homji and Mee [7,8]. The fog provides cooling when it evaporates in the air inlet duct of the gas turbine. This technique allows close to 100% effectiveness in terms of attaining 100% relative humidity at the gas turbine inlet and thereby gives the lowest temperature possible (the wet bulb temperature) without refrigeration. Direct high pressure inlet fogging can also be used to create a compressor intercooling effect by allowing excess fog into the compressor, thus boosting the power output considerably. In this paper, consideration is only made of *evaporative* fogging alone, with no discussion of fog intercooling being considered. A detailed parametric analysis of gas turbine response to fogging has been provided by Bhargava and Meher-Homji [9]. An application of fog intercool-

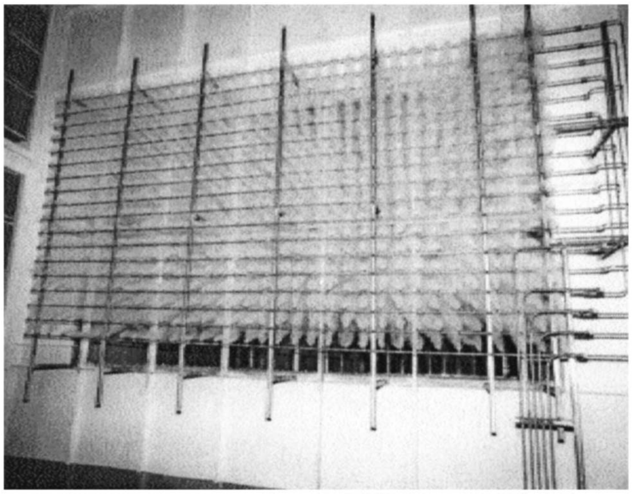


Fig. 3 High pressure fogging skid in operation for a heavy-duty gas turbine

ing on heavy-duty gas turbines is described by Ingistov [10]. A photograph showing a typical high pressure fogging skid is shown in Fig. 2.

Inlet fogging includes a series of high pressure reciprocating pumps providing demineralized water to an array of fogging nozzles located after the air filter elements. The nozzles create a large number of micron size droplets which evaporate cooling the inlet air to wet bulb conditions. A photo of a nozzle array fogging an inlet duct for a large frame gas turbine is shown in Fig. 3. A typical fog plume emanating from a single nozzle is shown in Fig. 4.

4 Climatic and Psychrometric Aspects of Inlet Fogging

4.1 Control of Inlet Fogging Systems and the Importance of Climatic Data. The control system of most inlet fogging systems incorporates a programmable logic controller (PLC), which is mounted on the high-pressure pump skid. Sensors are provided to measure ambient relative humidity and dry bulb temperature. Programming algorithms within the PLC use these measured parameters to compute the ambient wet bulb temperature and the wet bulb depression (i.e., the difference between the dry bulb and wet bulb temperature) to quantify and control the amount of evaporative cooling that is possible at the prevailing ambient conditions. The system turns on (or off) fog cooling stages to match the ability of the ambient air conditions to absorb water vapor.

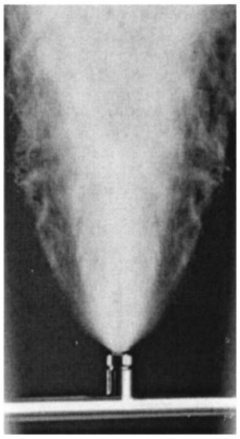


Fig. 4 Typical fog plume from a single fog nozzle

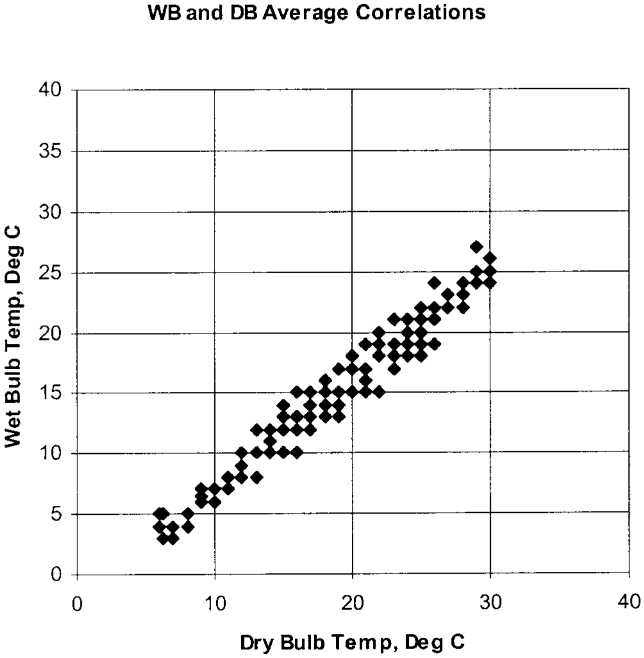


Fig. 5 Correlation of WB and DB temperatures—averaged data

The software would then be configured to adjust the amount of fog injected in proportion to the inlet air mass flow.

By choosing pump displacements (i.e., flow in gpm) it is possible to derive multiple cooling stages with the utilization of different pump combinations.

Obviously, the control of the skid is based on climatic conditions and so the overall utilization of the fogging system at any location, is a strong function of the climatic conditions. *It is this reason that makes an accurate understanding of the variations in climatic conditions an imperative*. McNeilly, [2] has addressed the area of climatic data for evaporative cooling.

4.2 Modeling of Climatic Data. There are numerous problems and traps when modeling climatic data-several of which derive from the concept of "averaging" of data. One example of this is using data such as shown in Fig. 5. This figure provides a correlation of dry bulb and wet bulb averages at a certain site. The graph shows that the linear behavior may lead one to conclude that at a dry bulb temperature of 25°C, the expected wet bulb is 20°C allowing a wet bulb depression of 5°C. (i.e., a measure of evaporative cooling potential). This is totally erroneous as the data was derived by taking the *average* WB temperature and the *average* DB temperature and plotting the curve. Consequently, the graph does *not* reflect *coincident* WB and DB conditions and will therefore indicate a much reduced cooling potential. It is advisable that the site's temperature profile for a full year of hourly data with the 20–30 year average wet and dry bulb coincident temperatures be considered in the analysis. These data can be used to generate evaporative cooling degree hour (ECDH) numbers for each hour of the year and allow a turbine operator to make a very detailed and accurate analysis of potential power gain from inlet fogging.

High relative humidity conditions do not occur with high dry bulb temperatures. A typical pattern of variation of dry bulb and wet bulb temperature over a day is depicted in Fig. 6. As can be seen, during the afternoon hours, there is a considerable difference between the wet bulb and dry bulb temperatures. It is this spread that allows the use of fog evaporative cooling.

A common mistake made by users is to take the reported high relative humidity and temperature for a given month and base the

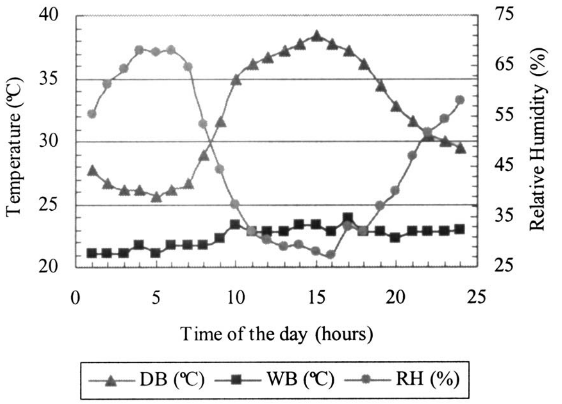


Fig. 6 Daily variation of dry bulb and wet bulb temperature

design on these. The problem is that the high relative humidity generally occurs time-coincident with the lowest temperature and the lowest relative humidity occurs with the highest temperature. This mistake results in the erroneous conclusion that very little evaporative cooling can be accomplished and has historically been the underlying cause of the maxim that evaporative cooling is not possible in "high humidity regions."

4.3 Fog Evaporative Cooling in High Humidity Regions. Even the most humid environments allow for up to 8°C (15°F) of evaporative cooling during the hotter part of the day. The term "relative humidity" refers to the moisture content in the air "relative" to what the air could hold at that temperature. In contrast "absolute humidity," is the absolute amount of water vapor in the air (normally expressed in unit mass of water vapor per unit mass of air).

The moisture-holding capacity of air depends on its temperature. Warmer air can hold more moisture than cooler air. Consequently, relative humidity is highest during the cool morning and evening hours and lowest in the hot afternoon hours. Since inlet air fogging systems cause a very small pressure drop in the inlet air stream, and are relatively inexpensive to install, they have been successfully applied in areas with very high summer time humidity such as the Texas Gulf Coast region or the state of Florida in the U.S.A. and in the other high humidity locations in the world. This is because during the hot hours, the coincident relative humidity is typically low.

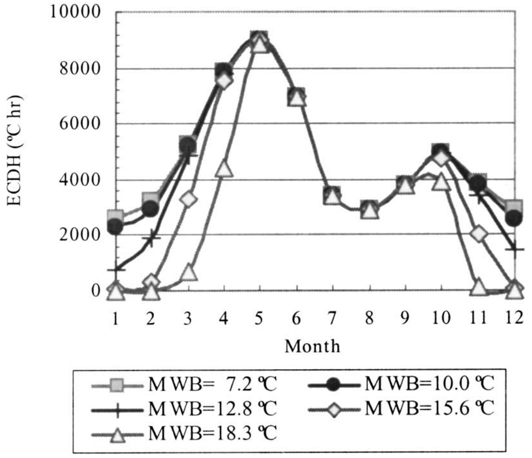


Fig. 7 Data for New Delhi, India. This provides the months of the year on the abscissa and the ECDH on the ordinate. The graph provided a month-by-month number of the ECDH, for a range of minimum wet bulb temperatures ranging from 18.3°C to 7.2°C. The "bimodal" pattern occurs here because after the hot months of April and May, temperatures drop due to the monsoons and then peak again in the months of September and October.

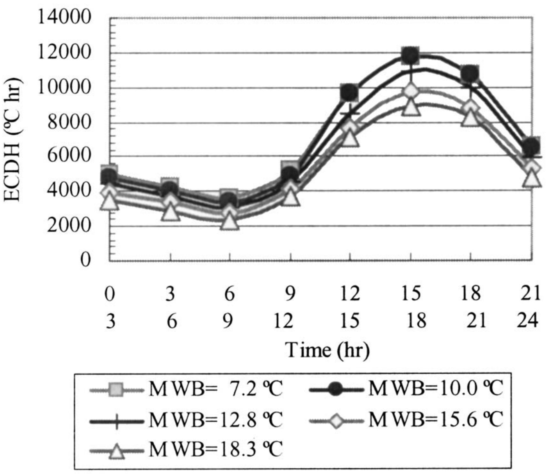


Fig. 8 Data for New Delhi, showing the duration of the day that the cooling potential is available. This graph shows that the cooling potential is predominantly clustered between 12:00–21:00 hours. Graphs such as this allow users to match evaporative cooling to peak demand needs.

5 Methodology and Analysis to Create an International Database for Evaporative Cooling Degree Hours

Data were obtained from a climatic database published by Airforce Combat Climatology Center. The goal of the analysis was to determine the equivalent cooling degree hours (ECDH) for a variety of locations worldwide. The ECDH is defined as a number that provides the total amount of cooling that can be derived for a given time period. The total ECDH is arrived at by summing the ECDHs derived for the 12 months at a location. Results are shown for approximately 106 cities in Table 1 in Appendix A. For example, the total ECDH for Athens, Greece of 29,900 ECDHs is derived by summing the numbers in that row, from January to December.

The database provides a wide range of information including dry bulb temperature values, and percentage of occurrence, wet bulb temperatures ranges and coincident dry bulb temperatures, and humidity ratios.

After data were collated from the above data files, a cross check was performed with ASHRE data. Finally, the data was placed in a spreadsheet and then a tabulation provided in the Appendix A was derived. The ECDH was chosen with a lower WBT limit of 55°F (12.8°C).[3] This was considered a very conservative number to avoid any possibility of inlet icing. Figures 7 through 12 show climatic data for New Delhi, India.

The number of ECDHs for which certain wet bulb depression ranges are available are shown in Table 2 of Appendix B. The ranges selected are in groups 0–5°C, 5–10°C, 10–15°C, 15–20°C, and >20°C. Due to rounding issues and temperature recording approaches, the summation of the ECDHs in this table, may differ up to 5% from the Table 1.

5.1 Use of the Equivalent Cooling Degree Hour (ECDH) Table in Appendix A. Any gas turbine operator can immediately see the potential for evaporative cooling per month in his or her location based on a long term historical database by using Table 1 in Appendix A. The results can be directly read off the tabulation providing the total ECDH and it is relatively easy to

---

[3] Figures 7–12 show the sensitivity of the results to the choice of minimum WBT.

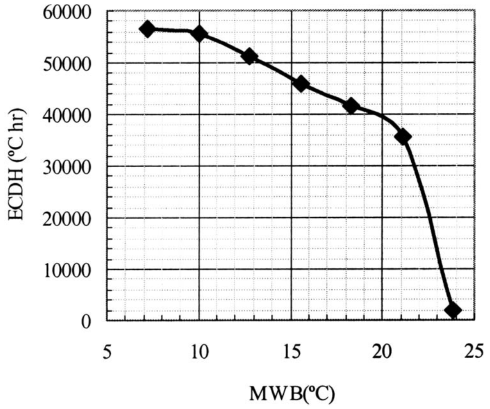


Fig. 9 Data for New Delhi showing the sensitivity of ECDH to the minimum wet bulb temperature. For example, if the minimum WBT is set to 15°C, then the value of the ECDH is approximately 47,000. If a more aggressive minimum temperature of 10°C is chosen, then the ECDH increase to 56,000.

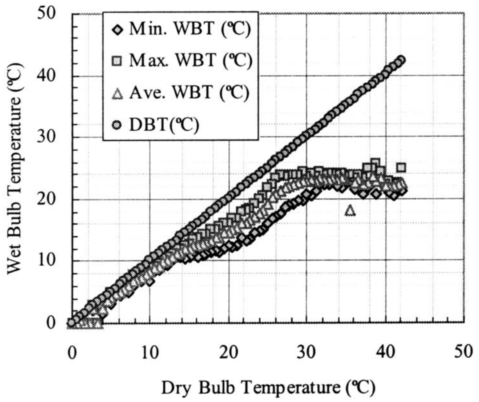


Fig. 10 Data for New Delhi, India. This shows the relationship between DBT and WBT. At a temperature approximately 40°C, a wet bulb depression of approximately 20°C is available.

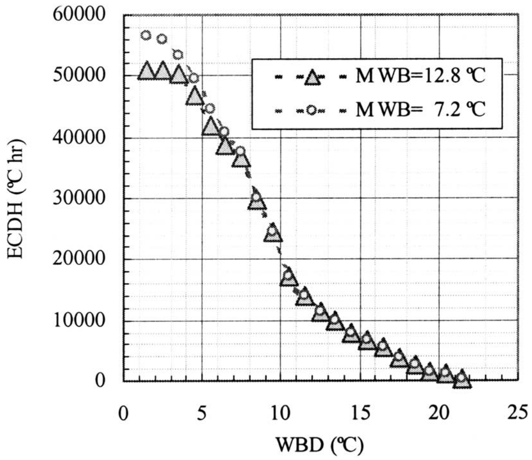


Fig. 11 Data for New Delhi, India. ECDH for a variety of wet bulb depressions. For example, at a wet bulb depression of 5°C, and a minimum wet bulb temperature of 7.2°C. At low WBDs the minimum west bulb temperature become important, as expected.

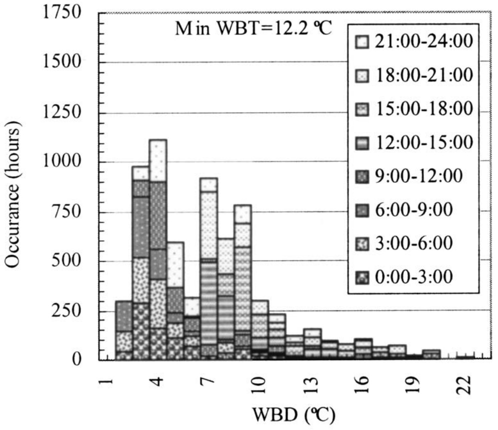


Fig. 12 Data for New Delhi, India. Graph showing the time frames during a day when the cooling hour potential exists for a range of web bulb depressions. At a WBD of 4°C, approximately 1125 hours exist and the time distribution of these hours is shown.

compute the kW hours of capacity available by the use of evaporative fogging. In order to do this, the ECDH number would be multiplied by the turbine specific kW/°C cooling number. This can be obtained from the gas turbine OEM's curves. In order to make the task easier, Table 3 in Appendix C provides a detailed tabulation of over 35 gas turbines (both 50 and 60 Hz operation) with runs made to examine the performance sensitivity to changes in ambient conditions. Gas turbines chosen range from 5–250 MW in output and the temperature effects on power are presented in terms of both power boost% per °C and also in terms of MW/°C. The use of this table will allow users to make an assessment of their gas turbines.

ECDH data can also be looked at more closely to account for differences in energy market values at different times of the year. For example, examination of data could provide an estimate of the revenue stream during the hot summer months alone. An economic evaluation can then be developed on a month-by-month basis knowing the site-specific economic criteria.

5.2 Practical Aspects. If the ECDH number is used to compute MW hr boost over the year, it is important to note that this would imply that fogging is employed whenever there is even a 1°C depression. In reality there may be a delay set in the control system to trigger the first stage of cooling and also the cooling degrees per stage, would have to be larger than the depression. Typical stage cooling is 1–1.7°C (2–3°F) Further, there is a practical limitation in that in the cooler months, it may be possible that freezing conditions may occur during the early morning and then during the day, an adequate temperature drop may exist. Some operators may drain the skid and hence may not make use of this period of evaporative cooling potential. If power is needed, however, then operators may elect to drain during the night and utilize the system during the day. Because of these factors it is impossible to make adjustments in the ECDH tabulation. Users may want to multiply the total number by a factor of 0.97[4] to account for these factors.

5.3 Lowest Temperature for Cooling. Table 1 (Appendix A) has been based on a minimum temperature of 12.8°C (55°F). This is a conservative number in terms of anti icing. Several OEMs publish a combination of relative humidity and temperature at which anti icing measures are turned on. With fogging applications where the ending relative humidity is close to 100%, temperatures as low as 10°C can be utilized.[5] However, to be on the

---

[4] This factor is an estimate, and may go as low as 0.94 in extreme cold locations.

[5] There are several considerations other than just calculating the intake temperature static depression caused by air acceleration to Mach numbers of 0.5 to 0.8. There is also some heating (although small—of the order of 1°C) due to the condensation that occurs and also due to heat transfer from the number 1 bearing, etc.

very conservative side, temperatures of 12.8°C have been considered.

6 Gas Turbine Simulation

In order to put the entire situation into perspective, a GTPRO simulation was made using a Frame 7121EA and a Frame 9351FA gas turbine in simple cycle configuration (fueled by natural gas) as a reference plant. Salient particulars of this gas turbine are when operating under varying wet bulb depressions of 8 and 12°C from the base temperature are presented in Table 4 (Appendix D).

7 Economic Criteria for Inlet Cooling

The specific decision to utilize inlet evaporative fogging technology is an economic one and the total project cost must be evaluated over the life cycle. Because of the varying economic situation in different parts of the country, no economic analysis is presented here. Dominating factors, which should be taken into account in doing a study, are

- climatic profile (discussed above).
- installed cost of the cooling system in terms of $/incremental power increase.
- amount of power gained by means of inlet air cooling. This should take into account parasitic power used, and the effect of increased inlet pressure drop. With fogging systems, the maximum parasitic power is in 50–80 kW for larger turbines when the maximum wet bulb depression has to be derived. The inlet pressure drop is almost nil due to the configuration and design of the nozzle array.
- fuel and demin water costs, and costs of incremental power, i.e., what benefit is attained by the power boost.
- projected O&M costs for the system.
- environmental impact.
- for cogeneration applications, the time of use electric rates and the power purchase agreement have to be carefully considered.
- potential impact on existing emission licenses.

Economic analysis for inlet cooling systems may be found in Utamura et al. [11], Ondryas [12], van Der Linden and Searles [13], and Guinn [14].

8 Closure

The paper has provided a tool to easily enable operators in worldwide locations to determine the degree of evaporative cooling potential in terms of evaporative cooling degree hours. The analysis and tabulations help users by distilling a huge amount of climatic data into a quick and easy to use format. Any gas turbine operator can use the data to get an idea as to the feasibility of the application of inlet fogging. Further, the data provided would reduce the effort that is needed to make an economic analysis of the potential of evaporative fog cooling.

Nomenclature

ASHRE = American Society of Heating and Refrigeration Engineers
CCPP = combined cycle power plant
DB = dry bulb temp, °C
ECDH = equivalent cooling degree hours, °C hrs (°F hrs)
EPC = engineering procurement and construction
HRSG = heat recovery steam generator
GPM = gallons/minute
LPM = liters/minute
NP = net power
WB = wet bulb temp, °C
WBD = wet bulb depression, °C
MWB = minimum wet bulb temp, °C
WG = water gauge

Appendix A

Table 1 Equivalent cooling degree hours (ECDH)–degree °C hr for selected cities

Available Yearly and Monthly ECDH in 106 cities worldwide
Based on Data From: Air Force Combat Climatology Center (AFCCC)
Min WBT=12.8 °C

	Ann.	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
Algeria, Dar El Beida	21331	245	367	834	1256	2021	2642	3270	3369	2942	2348	1407	630
Argentina, Buenos Aires	20370	3580	2659	2391	1473	763	274	261	467	894	1759	2540	3309
Australia, Adelaide	23134	4149	4046	3483	2113	887	170	11	86	394	1614	2600	3581
Australia, Melbourne	14004	2837	2729	2316	1178	358	27	4	21	236	758	1390	2151
Australia, Sydney	21490	2699	2419	2314	2200	1489	399	101	268	1391	2590	2661	2961
Austria, Salzburg	9564	0	0	20	85	1047	1878	2579	2416	1272	268	0	0
Bahamas, Nassan	27894	2112	2053	2512	2684	2547	2251	2550	2300	2242	2277	2147	2219
Bermuda, Bermuda	28194	1969	1540	1779	2324	2347	2161	2532	2871	2765	2789	2730	2385
Brazil, Belem	18543	1239	957	1057	1057	1385	1715	1871	1960	1906	1911	1859	1625
Brazil, Rio De Janeiro	29738	2883	2716	2587	2285	2096	2081	2265	2485	2382	2609	2649	2699
Burma, Rangoon	28702	3671	3879	4129	3942	2366	924	863	830	1111	1588	2257	3143
Canada, Halifax	6495	0	0	2	28	336	1413	1904	1752	871	168	20	1
Canada, Kamloops	15242	0	0	0	87	917	2992	4826	4538	1798	83	2	0
Canada, Montreal	10204	0	0	2	132	1121	2343	2797	2377	1154	251	28	0
Canada, Ottawa	10690	0	0	5	181	1107	2508	3043	2509	1080	234	23	0
Canada, Toronto	10699	0	0	20	232	1008	2360	2957	2478	1298	298	45	3
Canada, Vancouver	7496	0	0	0	65	552	1354	2132	2057	1158	172	7	0
Chile, Pudahuel	21158	4072	3361	2955	1580	498	91	44	265	595	1551	2563	3583
China, Beijing	17051	0	0	0	551	3847	4105	2623	2392	2775	758	0	0
China, Hong Kong	26221	1529	1299	1531	1735	1945	2106	2406	2447	2736	3300	3004	2184
China, Shanghai	12894	2	47	109	894	1991	1669	1889	1938	1911	1804	587	54
Colombia, Bogota	8394	729	876	1044	936	968	733	131	156	272	792	847	910
Costa Rica, San Jose	28680	3406	3297	3665	3277	2031	1555	1994	1911	1332	1448	1912	2851
Cuba, Havana	24076	1889	1874	2361	2534	2360	1955	2040	2001	1716	1724	1750	1873
Dominican, Caucedo	21988	1847	1826	2090	2085	1842	1880	1960	1956	1644	1618	1550	1687
Ecuador, Quito	1400	270	270	298	376	44	6	10	2	5	6	9	103
Egypt, Alexandria	27462	602	730	1336	2609	3061	3043	2916	3122	3266	3051	2412	1313
Egypt, Cairo	41696	77	325	2045	4908	6106	5990	5035	4719	4544	4207	2865	876
El Salvador, San Salvador	27213	3209	3200	3304	2922	1943	1490	1718	1546	1294	1625	2204	2757
France, Lyon	12904	0	0	33	216	1326	2213	3306	3295	1841	640	30	5
France, Marseille	18192	0	9	98	296	2158	3251	4212	3909	2635	1321	267	35
France, Paris	10512	0	0	23	181	960	1881	2756	2833	1439	391	47	1
Germany, Berlin	10300	0	0	0	114	1122	2004	2937	2866	1007	249	0	1
Germany, Frankfurt	10392	0	0	7	131	1108	1970	2880	2803	1250	238	4	0
Germany, Munich	8102	0	0	2	29	833	1495	2318	2216	1012	197	0	0
Germany, Stuttgart	8669	0	0	4	42	802	1587	2460	2424	1103	245	2	0
Greece, Athens	29900	64	88	223	1074	3254	4545	6022	6106	4669	2722	849	283
Greece/Crete, Khania	28161	49	55	213	1367	3451	4887	5192	4931	3848	2616	1190	364
Greece, Larissa	25179	5	5	145	752	2767	4746	5670	5176	3761	1845	288	19
Guatemala, Guatemala	25076	2131	2227	2735	2803	2448	1694	1968	1813	1479	1837	1913	2030
Honduras, Tegucigalpa	33254	2652	2979	4016	3971	3252	2408	2643	2643	2266	2075	2045	2304
Hungary, Budapest	13946	0	0	31	206	1779	2670	3493	3443	1901	422	1	0
India, Bombay	37751	4142	4094	4223	3305	3268	2257	1526	1506	1749	3108	4184	4390
India, Calcutta	29665	2596	2963	3959	3275	2757	1913	1549	1460	1555	2272	2726	2641
India, Delhi	51216	751	1900	4897	7806	8976	6984	3423	2898	3791	4931	3397	1462
India, Hyderabad	56761	4329	5193	7353	7841	8251	4465	2841	2431	2785	3577	3706	3987
India, Madras	36921	2485	2497	3030	3066	4177	4704	3928	3645	2899	2279	1937	2276
Iran, Tehran	41226	0	0	0	156	3106	8956	10968	10879	6779	381	1	0
Iraq, Baghdad	66801	41	202	1670	5027	8496	10760	11934	10943	9487	5937	2054	250
Ireland, Dublin	3730	0	0	0	11	125	662	1233	1077	486	117	19	1
Israel, Tel Aviv	26449.16	325	434	1436	2708	2689	2679	2846	3125	3272	3280	2659	997

Table 1 (continued).

	Ann.	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
Italy, Milano	13495	0	2	48	370	2012	2693	3129	2766	1816	646	14	0
Italy, Naples	15945	24	9	176	728	2167	2513	3050	2941	2185	1535	491	126
Italy, Venezia	12508	0	0	14	313	1869	2285	2802	2665	1785	731	44	1
Ivoiry Cost, Abidjan	16925	1692	1372	1565	1646	1578	1325	1099	906	948	1338	1694	1763
Jamaica, Kingston	35775	3033	2716	3089	3064	3062	3179	3437	3070	2704	2673	2677	3072
Japan, Fukuoka	16111	13	35	95	1120	2367	2190	2352	2856	2386	2098	555	43
Japan, Nagasaki	15405	15	32	108	1058	2106	1864	2054	2675	2443	2292	700	59
Japan, Osaka	18271	0	7	38	999	2681	2817	2897	3389	2811	2082	533	16
Japan, Tokyo	17307	7	14	40	965	2636	2446	2688	3062	2769	2266	389	23
Jordan, Amman	32838	3	8	122	964	3873	5931	6199	5980	5408	4001	329	19
Korea, Inchon	10895	0	0	0	153	1594	2050	1741	2062	2191	1058	46	0
Korea, Pusan	12494	0	8	8	585	2025	1815	1507	2041	2265	1922	307	10
Korea, Seoul	13518	0	0	0	439	2346	2684	2041	2329	2400	1234	45	0
Lebanon, Beirut	26717	278	476	1236	2397	2623	2815	3062	3076	3608	3552	2608	988
Libya, Tripoli	35984	244	620	1710	3030	4602	4993	5087	4954	4274	3582	2103	786
Malaysia, Kuala Lumpur	21914	1979	1913	1967	1659	1750	1912	1971	2131	1809	1686	1434	1703
Mexico, Mexico	18768	5	2	316	975	3287	3450	2693	2699	2498	2038	804	0
Morocco, Rabat	17497	500	681	1060	1351	1831	1892	2125	2137	2009	1783	1378	750
Netherlands, Amsterdam	6104	0	0	4	93	686	1110	1565	1675	759	201	9	0
Netherlands, H. V. Holland	5702	0	1	18	99	641	827	1382	1566	866	293	10	0
Netherlands, Soesterberg	7321	0	0	5	82	703	1277	2026	2074	915	233	6	0
Netherlands, Volkel	7672	0	0	5	130	830	1447	2050	2037	929	238	5	0
New zealand, Wellington	7288	1602	1384	1225	646	227	39	1	1	61	227	534	1341
Nicaragua, Managua	33984	3351	3571	4307	4382	3372	1974	2110	2220	1916	1766	2156	2858
Norway, Oslo	6328	0	0	0	11	333	1300	2482	2035	165	1	0	0
Pakistan, Karachi	44726	2707	3742	4785	4381	3885	3359	2748	2530	2976	5167	4792	3652
Panama, Tocumen	23012	2653	2773	3225	2777	1730	1320	1466	1466	1231	1166	1275	1931
Paraguay, Asuncion	33315	3506	2837	2995	2315	1854	1684	2139	2750	3013	3319	3409	3495
Peru, Lima	18664	1924	1736	1862	1505	1369	1351	1375	1328	1290	1492	1598	1834
Philippines, Manila	28836	2462	2567	3251	3378	3096	2295	2008	1860	1729	1908	2063	2219
Poland, Warsaw	8308	0	0	0	176	1130	1761	2167	2140	761	173	0	0
Portugal, Lisbon	19127	110	151	549	992	1855	2744	3542	3543	2990	1745	661	244
Puerto Rico, San Juan	30917	2567	2495	2983	2884	2624	2560	2560	2594	2402	2434	2314	2501
Russia, Moscow	6550	0	0	0	20	894	1657	2021	1505	448	4	0	0
Saudia Arabia, Dhahran	63262	1215	1756	3387	5398	7953	9430	9416	8466	6258	4823	3261	1899
Saudia Arabia, Riyadh	91256	412	543	2973	7999	11921	13428	14664	14426	12060	8183	3635	1013
Singapore, Singapore	17313	1595	1498	1708	1402	1439	1485	1546	1632	1377	1428	1040	1162
Spain, Barcelona	12546	4	28	80	485	1361	1933	2391	2337	2015	1365	485	62
Spain, Cordoba	29448	27	167	796	1433	2870	4147	5847	6085	4618	2457	782	218
Spain, Madrid	23415	0	1	111	380	1779	4087	6088	5823	3804	1211	129	3
Spain, Sevilla	31874	87	343	1127	1901	3404	4435	5879	5650	4539	2842	1239	428
Sweden, Stockholm	5871	0	0	0	22	394	1189	2269	1786	200	10	0	0
Switzerland, Geneva	10661	0	0	4	18	867	2001	3107	2930	1430	303	2	0
Taiwan, Taipei	20945	1031	902	1204	1579	1737	2006	2637	2511	2236	2049	1670	1384
Thailand, Bangkok	35935	3574	3103	3314	3347	2939	2770	2813	2768	2301	2425	2971	3611
Tunisia, Tunis	24166	222	283	721	1348	2687	3508	4216	3998	3078	2240	1324	541
Turkey, Ankara	15526	0	0	0	105	970	2634	4494	4691	2251	380	0	0
Turkey, Istanbul	16220	0	0	13	364	1776	2995	3413	3270	2696	1402	290	0
Turkey, Izmir	28045	32	69	313	1507	3325	4789	5647	5219	3912	2240	781	211
UK, Edinburgh	3196	0	0	0	15	83	494	1103	1073	340	86	2	0
UK, London Heathrow	8568	0	0	4	82	544	1553	2419	2409	1176	329	51	1
UK, London Gatwick	6755	0	0	2	42	428	1190	1979	1946	901	234	32	0
Uruguay, Carrasco	16244	2739	2093	2039	1313	661	261	247	448	722	1340	1861	2520
Vietnam, Ho Chi Minh	29772	3189	3142	3512	3172	2637	1951	1895	1903	1775	1750	2146	2702
Yugoslavia, Belgrade	14885	0	9	135	538	2082	2422	3118	3289	2324	897	66	6

Appendix B

Table 2 Availability of ECDH of a range of WBD (°C)

Min WBT=12.8 °C

	WBD				
Country, City	0-5.	5-10.	10-15.	15-20.	> 20.
Algeria, Dar El Beida	9666	9934	1382	173	2
Argentina, Buenos Aires	9021	10422	1019	5	0
Australia, Adelaide	4630	12048	6138	1373	62
Australia, Melbourne	3600	7172	2849	664	36
Australia, Sydney	13327	8522	832	176	8
Austria, Salzburg	2939	6522	418	5	0
Bahamas, Nassan	23427	4260	0	0	0
Bermuda, Bermuda	26897	1298	0	0	0
Brazil, Belem	13408	5694	2	0	0
Brazil, Rio De Janeiro	15727	11643	1903	17	0
Burma, Rangoon	13757	9962	4907	70	0
Canada, Halifax	4051	2101	14	0	0
Canada, Montreal	5203	4701	0	0	0
Canada, Ottawa	4409	5681	80	0	0
Canada, Toronto	4550	5417	59	0	0
Canada, Vancouver	6535	1272	1	0	0
Chile, Pudahuel	2885	12404	6149	68	0
China, Beijing	6005	10907	1207	54	0
China, Hong Kong	23809	2082	0	0	0
China, Shanghai	10938	1906	8	0	0
Costa Rica, San Jose	15063	13124	179	5	0
Cuba, Havana	16028	7606	3	0	0
Dominican, Caucedo	18687	2843	0	0	0
Egypt, Alexandria	14453	12240	380	116	2
Egypt, Cairo	7919	17423	14216	2411	159
El Salvador, San Salvador	13195	12808	808	0	0
France, Lyon	5155	6414	1085	40	0
France, Marseille	8234	9460	506	6	0
France, Paris	4874	5370	437	0	0
Germany, Berlin	3521	5987	708	10	0
Germany, Frankfurt	3872	5830	869	3	0
Germany, Munich	2856	5033	269	2	0
Germany, Stuttgart	3040	5164	412	2	0
Greece, Athens	6647	18772	4262	203	0
Greece/Crete, Khania	8732	15144	4039	519	8
Greece, Larissa	5569	12259	7481	846	52
Guatemala, Guatemala	12214	12540	524	2	0
Honduras, Tegucigalpa	12965	17437	2413	7	0
Hungary, Budapest	4152	8815	971	0	0
India, Bombay	18844	16982	1727	87	6
India, Calcutta	13033	15425	1048	114	0
India, Delhi	9218	24740	10229	5458	1139
India, Hyderabad	12782	23898	12841	6964	110
India, Madras	14881	17874	3495	205	0
Iran, Tehran	0	617	19861	19726	3789
Iraq, Baghdad	22	12781	23710	18502	12200
Ireland, Dublin	3467	219	0	0	0
Israel, Tel Aviv	16443	9579	423	197	3
Italy, Milano	6267	7226	34	2	0
Italy, Naples	8304	7485	286	2	0
Italy, Venezia	8209	4431	4	0	0

	WBD				
Country, City	0-5.	5-10.	10-15.	15-20.	> 20.
Ivory Cost, Abidjan	16704	472	17	5	0
Jamaica, Kingston	21025	14632	0	0	0
Japan, Fukuoka	10954	4992	43	0	0
Japan, Nagasaki	13161	1723	7	0	0
Japan, Osaka	9931	8428	104	0	0
Japan, Tokyo	11070	6530	49	0	0
Jordan, Amman	2345	16153	14609	1654	6
Korea, Inchon	9148	1786	17	0	0
Korea, Pusan	11968	550	0	0	0
Korea, Seoul	8318	5966	5	0	0
Lebanon, Beirut	18019	7648	411	85	11
Libya, Tripoli	8311	16585	8366	3148	105
Malaysia, Kuala Lumpur	12483	10429	19	0	0
Morocco, Rabat	14595	2255	564	82	6
Netherlands, Amsterdam	4558	1600	55	0	0
Netherlands, H. V. Holland	5074	772	64	6	0
Netherlands, Soesterberg	3802	3532	197	0	0
Netherlands, Volkel	4268	3361	197	0	0
New Zealand, Wellington	6561	443	0	0	0
Nicaragua, Managua	13228	17250	3112	0	0
Norway, Oslo	1385	4461	309	0	0
Pakistan, Karachi	13618	25195	5424	594	32
Panama, Tocumen	14664	8362	1	0	0
Paraguay, Asuncion	15685	15354	1931	29	0
Peru, Lima	17158	1069	0	0	0
Philippines, Manila	20330	8247	13	0	0
Poland, Warsaw	4030	4103	102	0	0
Portugal, Lisbon	9572	8067	1943	113	0
Puerto Rico, San Juan	20502	10015	8	0	0
Russia, Moscow	3038	3780	57	0	0
Saudia Arabia, Dhahran	7821	21525	18775	12128	2920
Saudia Arabia, Riyadh	0	1804	25904	32322	32566
Singapore, Singapore	14181	4068	0	0	0
Spain, Barcelona	11156	1534	8	0	0
Spain, Cordoba	5920	13155	7324	2678	12
Spain, Madrid	1989	11307	8956	807	0
Spain, Sevilla	7039	14087	8566	2040	4
Sweden, Stockholm	2066	3783	125	0	0
Switzerland, Geneva	3130	6738	621	0	0
Taiwan, Taipei	16937	3916	8	0	0
Thailand, Bangkok	15991	18972	568	0	0
Tunisia, Tunis	10683	11107	1808	160	29
Turkey, Ankara	695	7366	6648	594	2
Turkey, Istanbul	7732	8276	296	6	0
Turkey, Izmir	6887	14857	5659	159	0
UK, Edinburgh	2517	656	6	0	0
UK, London Heathrow	4329	3888	258	0	0
UK, London Gatwick	3778	2861	151	0	0
Uruguay, Carrasco	12239	3727	335	0	0
Vietnam, Ho Chi Minh	15677	13359	219	0	0
Yugoslavia, Belgrade	5552	8397	960	34	0

Appendix C

Table 3 Power boost of various gas turbines

plant Criteria:
DBT =32 °C R.H.=50% WBT=24 °C

Nomenclature: NF: No Fog
EC: Evaporative Cooling
NP:Net Power

Id	Gas Turbine	NP (NF) (MW)	NP (EC) (MW)	PB (%)	PB/°C	KW/°C
24	ABB GT 8	40.78	43.39	6.40	0.77	314.5
152	ABB GT 8C2	49.46	52.05	5.24	0.63	312.0
25	ABB GT 11N	69.86	73.78	5.61	0.68	472.3
110	ABB GT 11N2	97.24	103.35	6.28	0.76	736.1
41	ABB GT 13D2	86.77	91.82	5.82	0.70	608.4
70	ABB GT 13E2	139.00	147.51	6.12	0.74	1025.3
18	Aln 501KB5	3.02	3.25	7.62	0.92	27.7
21	Aln 571KA	4.44	4.81	8.33	1.00	44.6
129	Asig ASE40	2.62	2.77	5.73	0.69	18.1
151	Asig ASE50A	2.99	3.25	8.70	1.05	31.3
48	EGT Typhoon	3.24	3.63	12.04	1.45	47.0
145	EGT Typhoon	4.37	4.66	6.64	0.80	34.9
33	EGT Tornado	5.23	5.58	6.69	0.81	42.2
1	GE 5371 PA	22.15	23.86	7.72	0.93	206.0
133	GE 6561 B	34.29	36.42	6.21	0.75	256.6
135	GE 7241 FA	148.58	158.67	6.79	0.82	1215.7
174	GE 9351 FA	227.48	242.08	6.42	0.77	1758.8
113	GE 9391 G	246.85	258.54	4.74	0.57	1408.4
159	GE LM2500PE	18.79	19.88	5.80	0.70	131.3
118	GE LM6000PC	32.25	36.48	13.12	1.58	509.6
161	GE LM6000SPT	33.83	36.19	6.98	0.84	284.3
101	KWU V64.3A	60.04	63.19	5.25	0.63	379.5
43	KWU V94.2	125.96	133.46	5.95	0.72	903.6
154	Mtsb 701 F	228.76	241.82	5.71	0.69	1573.5
169	P+W ST6L-813	0.65	0.71	9.61	1.16	7.5
137	Sol Taurus	4.20	4.52	7.62	0.92	38.6
132	Sol Saturn	0.97	1.05	8.25	0.99	9.6
188	Sol Centaur	3.82	4.10	7.33	0.88	33.7
97	Sol Mars	9.12	9.74	6.80	0.82	74.7
189	Sol Titan	11.17	12.05	7.88	0.95	106.0
44	TP+M FT8	21.03	22.69	7.89	0.95	200.0
13	TP+M FT4C	24.21	26.48	9.38	1.13	273.5
47	W251 B12	39.94	42.89	7.39	0.89	355.4
77	W501 D5A	103.86	110.49	6.38	0.77	798.8
74	W701 F	197.08	209.30	6.20	0.75	1472.3

Appendix D

Table 4 Simulation run two heavy-duty gas turbines with inlet fogging

Tamb= 40°C Base case, Fuel is CH4 supplied at 25°C, No fuel compressor.
Fuel LHV 50047 kJ/kg; Inlet / Outlet DeltaP = 10 /12.5 mbar

	FRAME 7111EA (60 Hz)			FRAME 9351FA (50 Hz)		
GAS TURBINE PARAMETER	BASE CASE Tamb = 40°C	WBD = 8°C	WBD = 12°C	BASE CASE Tamb = 40°C	WBD = 8°C	WBD = 12°C
GT gross power [kW]	71,296	75,914	78,158	212,940	228,082	235,519
GT eff [%]	28.48	28.9	29.15	31.72	32.35	32.68
GT gross heat rate [kJ/kWh]	11,391	11,226	11,131	10,229	10,030	9,927
Fog Status	OFF	ON	ON	OFF	ON	ON
Compressor Pressure Ratio	11.2:1	11.6:1	11.9:1	13.1:1	13.6:1	13.8:1
Turbine Inlet Temperature [°C]	1104	1105	1107	1312	1313	1317
Exhaust Gas Temperature [°C]	554	549	546619	619	613	610
Compressor Inlet Mass flow rate [kg/sec]	260	269	274	575	594	606
Ambient Temperature [°C]	40	40	40	40	40	40
Compressor Inlet Temperature [°C]	40	32	28	40	32	28
Compressor Discharge Temperature [°C]	383	376	373	396	384	381
Compressor Disch. Pressure [Bar]	11.28	11.68	11.91	13.13	13.5	13.86
Fuel Flow Rate [Kg/sec]	4.5086	4.7299	4.828	12.089	12.6969	12.976
Axial Compressor Work [kW]	93,754	97,074	99,002	213,701	218,972	222,221
Turbine Section Work [kW]	166,895	174,885	197,084	431,281	451,889	462,675
Fog Water Flow, kg/sec	-	0.907	1.374	-	2.012	3.036

References

[1] Chaker, M., Meher-Homji, C. B., Mee, T., and Nicolson, A., 2001, ‘‘Inlet Fogging of Gas Turbine Engines—Detailed Climatic Analysis of Gas Turbine Evaporative Cooling Potential,’’ ASME Paper No. 2001-GT-526.

[2] McNeilly, D., 2000, ‘‘Application of Evaporative Coolers for Gas Turbine Power Plants,’’ ASME Paper No. 2000-GT-303.

[3] Kitchen, B. J., and Ebeling, J. A., 1995, ‘‘QUALIFYING Combustion Turbines for Inlet Air Cooling Capacity Enhancement,’’ ASME Paper No. 95-GT-266.

[4] Tawney, R., Pearson, C., and Brown, M, 2001, ‘‘Options to Maximize Power Output for Merchant Plants in Combined Cycle Applications,’’ ASME Paper No. 2001-GT-0409.

[5] Jones and Jacobs, 2000, ‘‘Considerations for Combined Cycle Performance Enhancement Options,’’ GE Publication GER-4200.

[6] Johnson, R. S., 1988, ‘‘The Theory and Operation of Evaporative Coolers for Industrial Gas Turbine Installations,’’ ASME Paper No. 88-GT-41.

[7] Meher-Homji, C. B., and Mee, T. R., 1999, ‘‘Gas Turbine Power Augmentation by Fogging of Inlet Air,’’ *Proceedings of the 28th Turbomachinery Symposium*, Turbomachinery Laboratory, Texas A&M University, Sept., Houston, TX.

[8] Meher-Homji, C. B., and Mee, T. R., 2000, ‘‘Inlet Fogging of Gas Turbine Engines—Part A: Theory Psychrometrics and Fog Generation and Part B: Practical Considerations, Control and O&M Aspects,’’ ASME Paper Nos. 2000-GT-307; 2000-GT-308.

[9] Bhargava, R., and Meher-Homji, C. B., 2002, ‘‘Parametric Analysis of Existing Gas Turbines With Inlet Evaporative and Overspray Fogging,’’ ASME Paper No. 2002-GT-30560.

[10] Ingistov, S., 2000, ‘‘Fog System Performance in Power Augmentation of Heavy Duty Power Generating Gas Turbines GE Frame 7EA,’’ ASME Paper No. 2000-GT-305.

[11] Utamura, M., Ishikawa, A., Nishimura, Y., and Ando, N., 1996, ‘‘Economics of Gas Turbine Inlet Air Cooling System for Power Enhancement,’’ ASME Paper No. 96-GT-515.

[12] Ondryas, I. S., ‘‘Options in Gas Turbine Power Augmentation Using Inlet Air Chilling,’’ ASME Paper No. 90-GT-250.

[13] Van Der Linden, S., and Searles, D. E., 1996, ‘‘Inlet Conditioning Enhances Performance of Modern Combined Cycle Plants for Cost-Effective Power Generation,’’ ASME Paper No. 96-GT-298.

[14] Guinn, G. R., 1993, ‘‘Evaluation of Combustion Gas Turbine Inlet Air Precooling for Time Varying Annual Climatic Conditions,’’ ASME Gogen-Turbo 1993, Bournemouth, UK, Sept. 21–23, IGTI, Atlanta, **8**.

Mustapha A. Chaker
Director, Research and Development,
Mee Industries, Inc.,
204 West Pomona Avenue,
Monrovia, CA 91016-4526

Cyrus B. Meher-Homji
Principal Engineer,
Turbomachinery Group,
Bechtel Corporation,
3000 Post Oak Blvd., MS 73,
Houston, TX 77056-6503

Thomas Mee III
Chairman and CEO,
Mee Industries, Inc.,
204 West Pomona Avenue,
Monrovia, CA 91016-4526

Inlet Fogging of Gas Turbine Engines: Experimental and Analytical Investigations on Impaction Pin Fog Nozzle Behavior

The inlet fogging of gas turbine engines for power augmentation has seen increasing application over the past decade. This paper provides the results of extensive experimental and theoretical studies conducted on impaction pin fog nozzles. It covers the important area of the fog plume pattern of impaction pin nozzles and examines fog-plume uniformity. The subject of sprinkle (large droplet formation) from the nozzles is also examined in detail and is shown to be nonsignificant. The effect, on evaporation rate, of ambient climatic conditions and the location of the fog nozzle with respect to the gas turbine inlet duct has been analytically and experimentally analyzed. An analytical model is used to study the evaporation dynamics of fog droplets injected in the inlet ducts. The model is validated experimentally in a wind tunnel. [DOI: 10.1115/1.1808429]

1 Introduction and Background

Gas turbine output is significantly impacted by the temperature of the ambient air, with power output dropping by 0.54% to 0.90% for every 1°C (1.8°F) rise in ambient temperature. This loss in output presents a significant problem to utilities, cogenerators, and merchant power plants, particularly when electric demands are high during the hot months. Inlet fogging is a popular technique for boosting the hot-weather performance of gas turbines and has been extensively implemented over the past decade. Inlet fogging fits the niche of low-cost power augmentation and has consequently become very popular.

The concept is a simple one in which a direct evaporation effect is derived by the use of fog generated by high-pressure pumps and atomizing nozzles installed in the inlet duct downstream of the inlet air filters. The fog evaporates in the inlet duct and cools the air down to the wet bulb temperature.

It is estimated that more than 700 gas turbines have been fitted with fogging systems, including many modern class F gas turbines. In spite of this proliferation of inlet fogging systems, there is not much technical literature available covering the thermodynamics, physics, and engineering principles that govern the fogging process. The first rigorous and detailed treatment was made by Chaker et al. [1–3] and this paper builds on that foundation.

Fog intercooling,[1] which has been applied from the early days of gas turbine and jet engine technology is a technique that consists of spraying more fog than will evaporate under the given ambient temperature and humidity conditions so that nonevaporated liquid water droplets enter the compressor. The desired quantum of unevaporated fog is carried with the air stream into the compressor, where it evaporates and produces an intercooling effect. The resulting reduction in the work of compression can give a significant additional power boost and an improved heat rate.

A review of the basic principles of fogging technology can be found in Meher-Homji and Mee [4,5]. Early papers on fog intercooling and wet compression started to appear in the late 1940s including Kleinschmidt[2] [6], and Wilcox and Trout [7]. Other references include Hill [8], Arsen'ev and Berkovich [9], Nolan and Twombly[3] [10], and Utamura et al. [11].

Most inlet fogging systems currently in operation employ one of two different types of fogging nozzles; swirl jet nozzles or impaction pin nozzles. This paper studies the behavior of impaction pin nozzles designed by Mee Industries as shown in Fig. 1 and gives some commentary on swirl jet type nozzles as well. There are several misconceptions regarding the behavior of impaction pin nozzles and this paper makes a systematic study of several of them.

Topics covered in this paper include:

- *Effects of airflow velocity and ambient relative humidity on droplet size*. Fogging systems have to operate at varying humidity conditions but, at this time, there has been no published systematic study of the effect of ambient conditions on fogging system performance.
- *Effects of fog plume shape and nonuniformity caused by displacement of the impaction pin*. In this paper we have made detailed measurements of the effect of partial pin displacement on droplet size and examined the behavior of the droplets in multiple locations in the spray plume. A discussion is also made of the forces required for movement of the impaction pin.
- *Large droplet formation*. A systematic study has been done of what we call the *sprinkle effect*. This effect can occur when the spray plume strikes the support side of the impaction pin. Water can accumulate on the pin and form large droplets, which are then stripped off and enter the airflow. The research presented here shows that the sprinkle effect occurs, with Mee impaction pin nozzles, only at lower than normal operating pressures, and that sprinkle occurs with both swirl jet

---

[1] Also known as overspray or wet compression.

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received October 2002; final manuscript received March 2003. Assoc. Editor: H. R. Simmons. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Atlanta, GA, June 16–19, 2003, Paper No. 2003-GT-38801.

[2] This paper coined the term "wet compression."

[3] Nolan and Twombly covered an application of fog intercooling on GE Frame 5 engines.

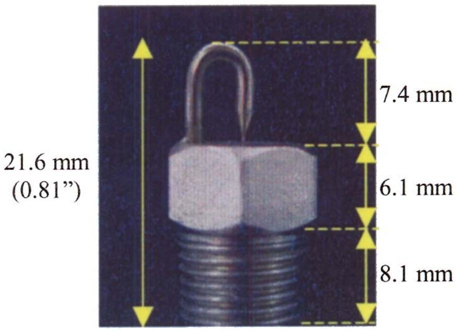


Fig. 1 Impaction-pin nozzle

and impaction pin nozzles. But, in both cases it has an insignificant effect on the performance of the fogging system or on the compressor blades.

- *Heat and mass transfer model*. A detailed droplet heat and mass transfer model was developed by Chaker [1] and is extended in this paper to include application to the whole range of droplets found in a typical inlet fogging spray. The results are provided for a range of ambient psychrometric conditions, and graphical results are provided to assist gas turbine users in understanding the significance of droplet size on fog system performance. This model allows the calculation of evaporative efficiency and predicts the size of non-evaporated droplets that may enter the compressor. It also quantifies the reduction in evaporative efficiency that naturally occurs with an increase in ambient relative humidity. While this phenomenon is intuitively clear, the model and curves provide greater insight and a means for making practical calculations.
- *Effects of fog nozzle location on fog system performance*. The position of the nozzle array in the inlet duct is an important design consideration as it affects both the size of the initial droplet (due to the effect of airflow) and the residence time for the droplets in the duct prior to reaching the compressor. This subject is discussed in detail including an example of a typical installation.

2 Experimental Setup

In order to measure droplet sizes in conditions similar to those found in gas turbine inlet ducts, a variable speed wind tunnel was used, as shown in Fig. 2. A brief description of the wind tunnel is given below, and more information regarding its configuration and instrumentation may be found in Chaker [1,3]. Fog is generated in the wind tunnel by forcing high-pressure, filtered and deionized water through the small orifice nozzles. In the experimental setup, a variable-speed-drive, positive-displacement, ceramic-plunger pump is used to generate water pressures up to 207 bar (3000 psi).

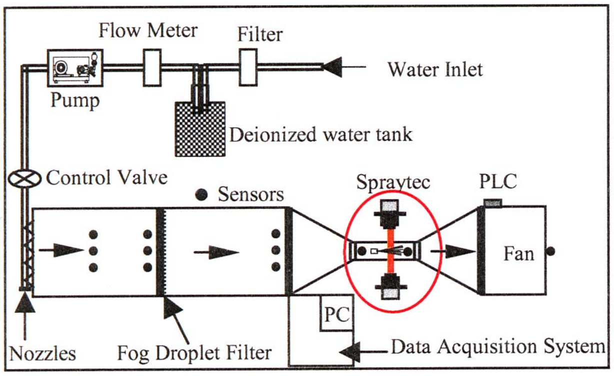


Fig. 3 Wind tunnel experimental setup

Droplet size measurement was done with nozzles located in the constricted section of the duct, where the highest airflow velocities are attainable. Measurements were taken at different locations in the spray plume using a Malvern Spraytec RTS5114 laser particle analyzer, as shown in Fig. 3. The Malvern system is based on a laser diffraction technique [12]. This is a spatial sampling system; consequently, it allows the sampling of a large number of droplets instantaneously with a frequency up to 2500 Hz.

A nozzle manifold, installed at the inlet of the duct, was turned off when measurements were taken at ambient conditions. For measurements with a saturated airflow these nozzles were turned on (Fig. 4) and a fog droplet filter was used, as shown in Fig. 3, to remove any unevaporated droplets.

Droplet size measurements, as shown in Fig. 5, were taken in the constricted portion of the wind tunnel, at different locations in the spray plume, and at different airflow velocities.

3 Effect of Air Velocity and Humidity on Droplet Size

It is important to understand the effects of airflow velocity and humidity on the atomization process because the nozzle manifolds can be installed at different locations in the inlet duct (with differing air velocities) and fog systems are operated under different ambient conditions.

In order to determine the effect of velocity, independent of air temperature and relative humidity (RH), studies were done result-

Fig. 2 Experimental wind tunnel, 10.5 m (34.5 ft.) long and capable of velocities up to 25 m/s (4900 ft/min). Used to study droplet kinetics and thermodynamics under conditions similar to gas turbine inlet air ducts.

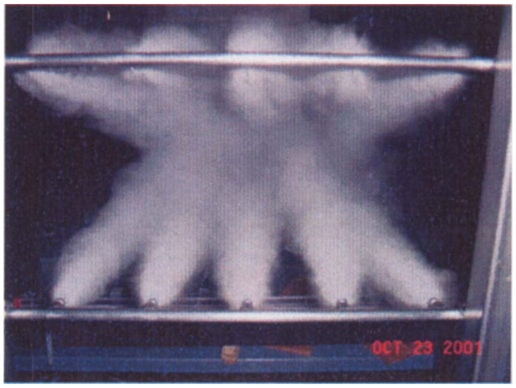


Fig. 4 Fog nozzles manifold used to create saturated air in the wind tunnel. The use of a fog droplet filter makes it possible to achieve an airflow with very close to 100% RH.

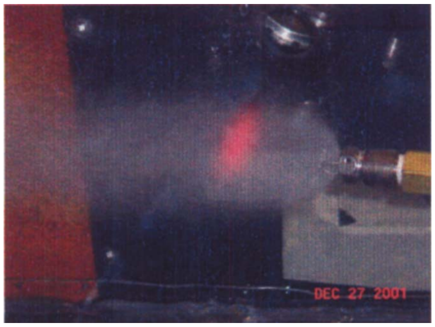


Fig. 5 Droplet size measurement in the wind tunnel

ing in Fig. 6, which shows the droplet sizes in the center of the plume and the weighted (for water flow) average across the plume. Droplet size measurements were taken at an axial distance of 7.6 cm (3 in.) from the nozzle. The airflow velocity varies from 0 to 15 m/s (0 to 3000 ft/min).

From Fig. 6, it can be clearly seen that there is a significant decrease in the measured droplet size when the velocity is increased from 0 and 5 m/s (984 ft/min). This is due to the fact that the droplet population near the nozzle is very dense and many collisions occur. Collisions result in droplet agglomeration. As the air velocity is increased, droplets of different sizes (with inherently different penetration velocities) are separated into different flow paths and droplet collisions are markedly decreased. At a higher airflow velocity, the smaller sized droplets react quickly to the airflow and take different trajectories as compared to the larger droplets.

To give an idea of the different response times, the largest droplets, of the order of 50 microns, have a response time of less than 8 milliseconds, while droplets of 10 microns have a response time of only about 0.3 ms. Therefore, when the difference between droplet velocity and airflow velocity near the nozzle orifice is around 10 m/s (1970 ft/min.) a droplet of 50 microns needs 8 cm (3.15 in.) to adjust its velocity to the airflow velocity, while a droplet of 10 microns needs only 3 mm (0.12 in.).

At higher air velocities, the center-of-plume measurements become much smaller than the average measurements because more and more of the smaller droplets are pushed towards the center of the plume, due to their faster response time.

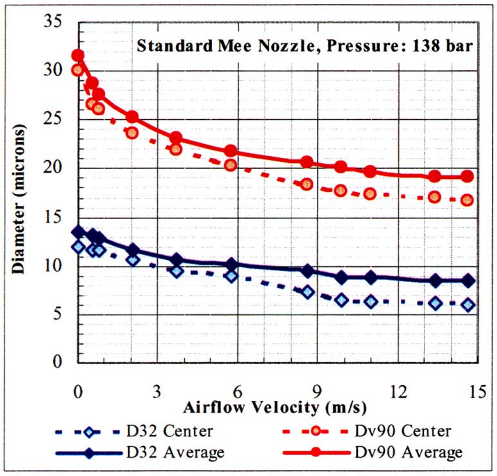


Fig. 6 Variation of the droplets size as a function of airflow velocity; measurement were taken at 30°C and 40% RH and at 7.6 cm from the nozzle orifice

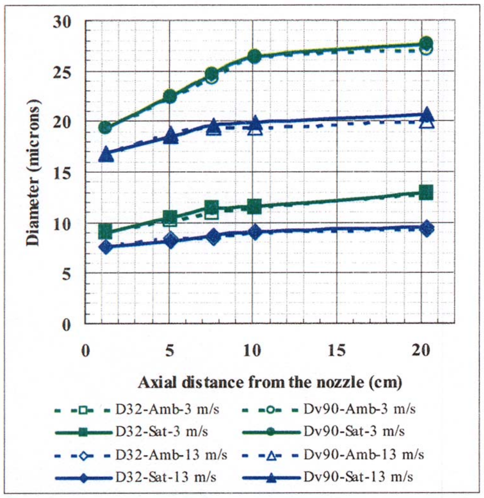


Fig. 7 Effect of ambient humidity on droplet size at different axial distances from the nozzle at 138 bar (2000 psi) pressure

Droplet size increases as measurements are taken farther from the orifice (to a limit) because of droplet collisions and coalescence. This increase in droplet size with increasing axial distance (Z) is shown in Fig. 7. Note that the effect of higher air velocity, i.e., decreased droplet size due to fewer collisions, is more pronounced close to the exit of the nozzle and diminishes as the axial distance increases until it disappears, for the nozzle tested, at around 20 cm (7.8 in.) downstream from the nozzle exit. Nozzles which produce inherently larger droplets, or which have higher flow rates, can have coalescence effects that continue as much as 50 cm (20 in.) downstream.

In order to experimentally quantify the effect of rapid small droplet evaporation, measurements were taken at two airflow velocities, 3 and 13 m/s (590 ft/min and 2560 ft/min), and for two sets of air conditions; 30°C (86°F) with 40% RH (defined as ambient in the chart) and 20°C (68°F) with 100% RH (defined as saturated).

Figure 7 shows that droplet size near the orifice is not much affected by the psychrometric properties of the air, while the velocity of the air has a dramatic effect. There is only a very small difference in the droplet size measured in cool, saturated air (the solid lines) where evaporation cannot occur, and the size measured in hot, dry air (the solid lines) where evaporation could occur. On the other hand, there is a big difference between droplet sizes at low air velocity (the dark violet lines) and at high air velocity (the green lines.)

Note that the two sets of lines high on the chart show the Dv90 diameters, while the two sets of lines lower on the chart give the D32 diameters. The reason for the difference in the two droplet sizes can be understood from the definitions given in the Appendix. In studying the curves, one can see that the dashed lines, representing droplet size at nonsaturated conditions, start to fall away from the solid lines, representing droplet size with dry air conditions, as the fog moves away from the nozzle orifice. This shows the very small effect of evaporation.

4 Effect of Plume Shape on Droplet Size

4.1 Pin Dislocation and Droplet Size. Concern has been voiced that the dislocation of the impaction pin could cause deterioration in the plume shape and that this might result in an increase in the size of droplets produced by the nozzles. First, it should be noted that dislocation of the impaction pin nozzle is not easily accomplished due to its structural strength. The pin can be deflected only by striking it a blow with a hard object or by the

Fig. 8 Distorted spray plume caused by intentional bending of the impaction pin, 138 bar (2000 psi) operating pressure

use of pliers. It is not possible to distort the pins by hand.[4] During installation, the nozzles are protected with plastic caps that are removed only after all work of installation has been finished and the system is ready for operation. The nozzles are unlikely to be damaged thereafter, as they are located inside the inlet duct.

In order to understand and quantify the effect of pin dislocation, a nozzle's impaction pin was physically bent using pliers so that the nozzle would produce a distorted spray plume, as shown in Fig. 8. The photograph clearly shows that the fog in the upper part of the plume is denser than in the lower sector. Detailed measurements of droplet size were then taken in the distorted plume. The nozzle was rotated, in increments of 45 deg, so as to characterize the droplet size in the plume in all directions.

Droplet tests were done at two airflow velocities and at 1.3 cm (0.5 in.) from the nozzle orifice. Results given in Fig. 9 show that even though the plume looks distorted, the droplet diameter sizes (D32 and Dv90) are remarkably constant. Results for measurements taken at 7.6 cm (3 in.) away from the nozzle orifice, as given in Fig. 10, also show minimal scatter, about 2 microns. These tests indicate that as long as the impaction pin is above the orifice, even in a noncentered position, the nozzle characteristics do not change significantly. Furthermore, nozzles with highly distorted spray plumes are easily found and replaced during system startup or periodic inspections.

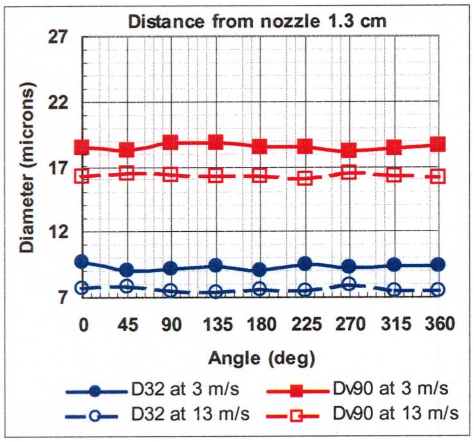


Fig. 9 Droplet size at different rotational angles for the distorted-plume nozzle measured at 1.3 cm (0.5 in.) from the nozzle. Operating pressure is 138 bar (2000 psi).

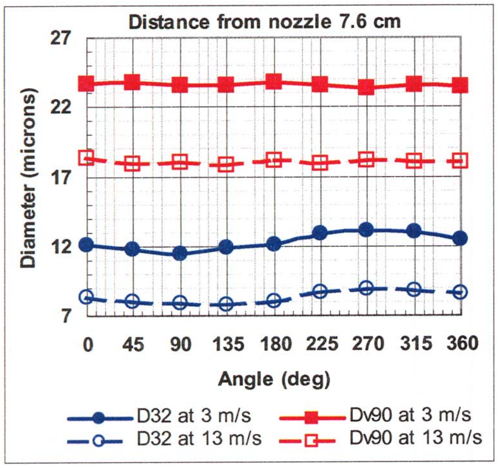


Fig. 10 Droplet size at different rotational angles for the distorted-plume nozzle measured at 7.6 cm (3 in.) from the nozzle. Operating pressure is 138 bar (2000 psi).

By comparing droplet sizes of the undamaged nozzle shown in Fig. 7 and the distorted plume nozzle shown in Fig. 9 and Fig. 10, one can see that at both distances from the nozzle orifice the measured diameters are essentially the same.

4.2 Creation of the Nozzle Spray Plume. The shape of the fog plume created by the impaction pin nozzle at different operating pressures was studied in this set of experiments. Close-up, high-speed photographs of the nozzle spray plumes (Figs. 12 and 13) were taken in order to better understand plume formation and atomization process. With the impaction-pin nozzle, water is forced through a short, smooth orifice and hits the impaction pin at a velocity that depends on the applied pressure. At 138 bar (2000 psi) the water jet exits the orifice at a velocity of about 160 m/s (525 ft/s). A properly designed impaction pin nozzle splits the water jet when it impinges on the sharp tip of the pin, and a thin, conically shaped sheet of water is formed. (A poorly designed impaction pin does not form a sheet and produces larger droplets.) The water sheet thins as it expands, then breaks apart into small droplets. Breakup occurs when the aerodynamic forces, which result from turbulence caused by the extremely high velocity of the sheet, overcome the surface tension of the water.

At low operating pressure, surface tension causes the conical sheet to begin to close up. This phenomenon can be seen in Fig. 12(*b*), where the operating pressure is 34 bar (500 psi). At very low pressures, the sheet closes back on itself and forms a hollow, bulb-shaped structure. As pressure is increased, the momentum of the sheet forces it to open up until it thins to a point where surface tension eventually rips it apart. This results in the formation of fingers or ligaments of water, which then break apart to form small droplets. Finger or ligament formation can be clearly seen in Fig. 12(*a*). At pressures higher than about 103 bar (1500 psi), Fig. 12(*c*) for instance, finger or ligament formation is no longer visible.

4.3 Droplet Formation and the Weber Number. The Weber number is a dimensionless number that is helpful for understanding and predicting sheet or droplet breakup caused by aerodynamic forces. The Weber number is the ratio of aerodynamic forces to surface tension forces, and is given by the equation

$$\mathrm{We} = \frac{\rho_{\mathrm{a}} \mathrm{V}_{\mathrm{rel}}^2 \mathrm{D}_{\mathrm{d}}}{\gamma_{\mathrm{w}}}.$$

When the Weber number exceeds about 13 [13], the forces of aerodynamic turbulence overcome the force of water tension and breakup occurs. In theory, the mean droplet size produced is esti-

---

[4] An axial force of about 200 lbs would result in breakage. Fatigue tests indicate that the pin would have to be distorted back and forth by 120 deg 8 times before failure.

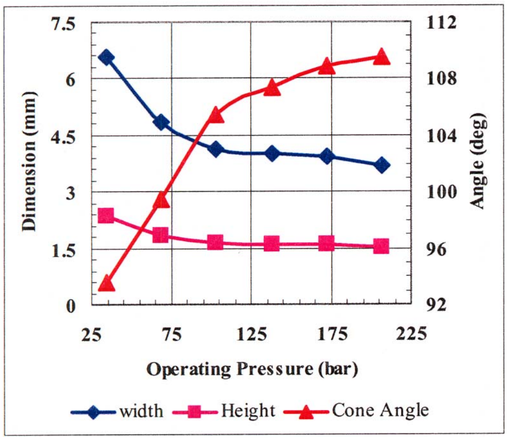


Fig. 11 Cone angle, height, and width of the conical water sheet at the point of atomization

mated to be proportional to the square root of the sheet thickness at the point of breakup [14]. For a given nozzle, increasing the operating pressure increases the sheet velocity so that the point at which sheet breakup occurs moves closer to the tip of the pin. This effect is shown in Fig. 11, which was plotted using measurements taken from the photos in Figs. 12 and 13.

The conical sheet formed by the impaction-pin nozzle expands and slows as it moves through the air. The thickness of the sheet is a function of the orifice and pin geometry and, in the absence of drag, would be proportional to the inverse square of the distance from the point of the impaction pin. In the real world, drag acts to increase sheet thickness, or keep it about the same, by causing the sheet flow to back up on itself.

As mentioned above, a thinner sheet will produce smaller droplets, all other things being equal. When operating pressure is increased the point of atomization moves closer to the orifice but the cone angle also increases. At this new point of atomization the sheet is probably actually thinner so that smaller droplets are produced. This agrees with our empirical data, which show smaller droplets and higher flow rates for increasing operating pressure.

The Weber number seems to approximately apply to the sheet breakup for the nozzle being discussed. If we assume demineralized water at normal ambient temperature and standard atmospheric air, then the term ρ_a/γ_w is equal to about 17.36. If we assume the sheet velocity is about 100 m/s (328 ft/s) at the point of atomization (i.e., we assume that it has slowed considerably from its initial velocity of 160 m/s) and assume a sheet thickness as 75 microns $[75\times10^{-6}\ \mathrm{m}]$ (i.e., assume it has not thinned), then we get a Weber number of 13. In fact, any combination of sheet velocity and thickness, the product of which is equal to 0.75, will yield a Weber number of about 13. But, since we know the velocity didn't *increase* and we know the sheet thickness is of the order of 75 microns, we can approximately determine the Weber number.

The Sauter mean diameter (SMD or D32) of a fog spray is reported to be approximately equal to the square root of the thickness of the sheet at breakup. Using the laser particle analyzer described above, the SMD for this nozzle was found to be about 9 microns. In the above approximation we assumed a sheet thickness of 75 microns, at 138 bar (2000 psi). The square root of 75 is 8.6, which shows a surprising degree of agreement with the measured values. Confirming these approximations will be the subject of future research at Mee Industries.

5 Nozzle Sprinkle and Pin Shading Effects

When operating pressure is low, the impaction-pin nozzle may exhibit large droplet formation, which we call sprinkle. A detailed study of large-droplet formation was conducted and results are given here. Sprinkle occurs when the conical water sheet contacts the wider portion of the impaction pin. Airflow or gravity eventually strips these very large droplets off the pin, and they are broken up by collision with the high velocity spray plume.

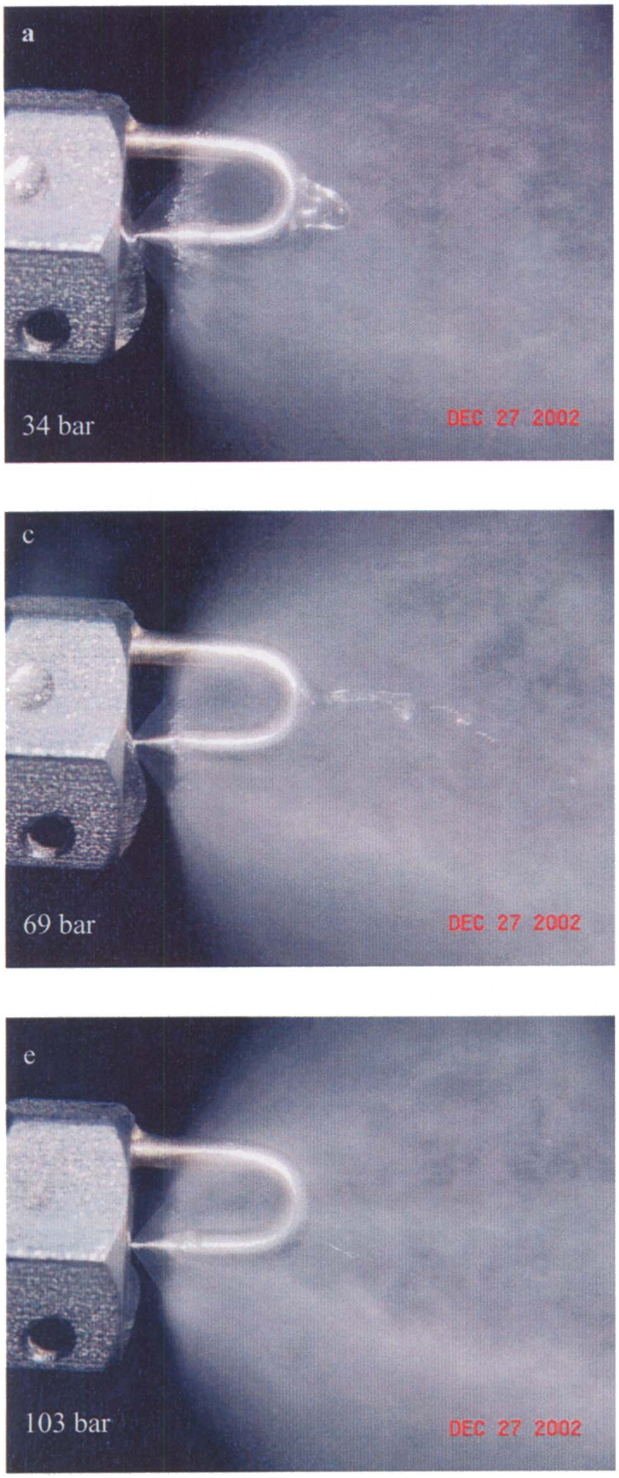


Fig. 12 Sprinkle effect from a standard Mee nozzle at operating pressures from 34 to 103 bar (500 to 1500 psi). The views are facing the impaction pin. Note how sprinkle lessens with increasing operating pressure.

Sprinkle formation can be seen in the photos in Fig. 12, which shows spray plumes generated at operating pressures ranging from 34 to 103 bar (500 to 1500 psi). One can see that the size of drops

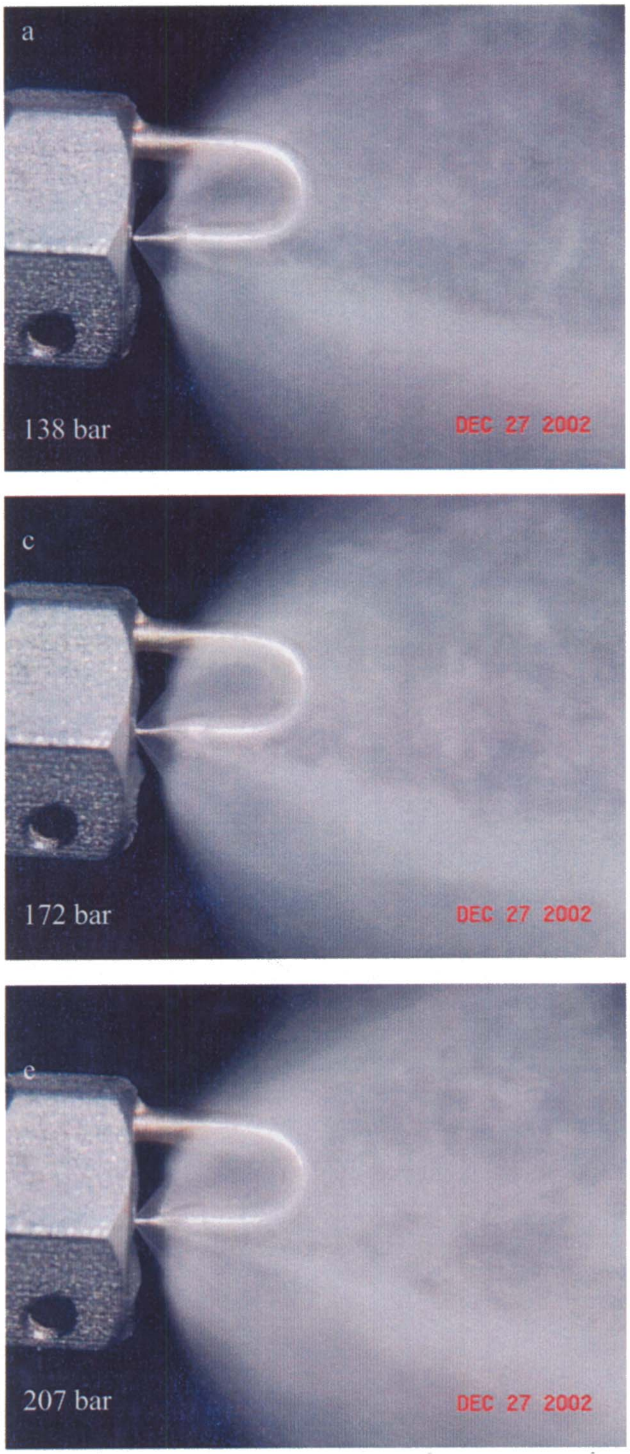


Fig. 13 Standard Mee nozzle spray plumes at operating pressures from 138 to 207 bar (2000 to 3000 psi). Note the absence of larger droplets (sprinkle) at these operating pressures. One can also clearly see that the atomization process begins well before the conical water sheet contacts the support side of the impaction pin.

emitted from the pin decreases significantly and becomes almost invisible at the higher pressures. Looking at Fig. 13, we see that, at 138 bar (2000 psi), sprinkle formation has completely disappeared.

The quantity of water emitted from the pin as larger droplets (sprinkle) was estimated at an operating pressure of 34 bar (500 psi) by estimating the size of the droplets and counting the rate of formation. It was observed that droplets of about 2 mm were produced at the rate of about 1 per second, which equates to about 0.3% of the total mass flow of the nozzle. The 2 mm droplets shatter when they are ejected into the vigorous spray plume, and smaller droplets are formed. The laser particle analyzer is capable of measuring droplets up to 400 microns, with its current lens configuration, but since the mass flow is so low the large-droplet data have only a negligible effect on the measured droplet size.

During nozzle testing in the wind tunnel it was noticed that swirl jet nozzles also produce sprinkle. The lower exit velocity of the swirl jet nozzles results in some very small droplets being recirculated back to the face of the nozzle, where they agglomerate and eventually form a large droplet. When this droplet reaches a critical size it falls off the nozzle or is suctioned back into the spray plume, where it is shattered into smaller droplets. This mode of sprinkle occurs even at the higher operating pressures. No attempt was made to quantify the amount of sprinkle from a swirl jet nozzle, but it is probably also a statistically insignificant amount. It was also noted that impaction-pin nozzles other than the Mee nozzle, which do not have good sheet formation, produce sprinkle even at high operating pressures.

In the typical operating range for impaction-pin nozzles, which is 140 to 210 bar (2000 to 3000 psi), the sprinkle effect either goes away completely or is too small to be visible. This is because at higher operating pressures the point of atomization occurs before the water sheet reaches the pin and the fog droplets are small enough to follow the flow lines around the pin.

Our conclusion is that sprinkle from impaction-pin nozzles is either nonexistent or insignificant at normal operating pressures.

6 Droplet Heat and Mass Transfer Model

Before trying to understand the complex properties and behavior of the whole mass of fog droplets emitted from the nozzle, it is helpful to study the properties and behavior of individual droplets. This understanding can then be extended to study the interaction between a single droplet and the other droplets and their interaction with the carrying phase (air). The study of an isolated droplet is first treated using the classical assumption of spherical symmetry for both the liquid droplet and the surrounding air. The behavior of a single droplet injected into the airflow will first be analyzed, then a study of the effect of the velocity on droplet thermodynamics and trajectory will be done.

A numerical model of droplet behavior was developed (Chaker et al. [1]). This model works in iterative manner and provides the transient behavior of fog droplets in terms of the droplet diameter, change in relative humidity, and temperature and time to attain saturation. The iterations stop ($I=I_{max}$) when the air in the volume around the droplet within active radius (RA) becomes saturated or when the droplet evaporates completely. The process is shown in the flow chart of Fig. 14.

The model makes a quantitative analysis of fog droplet behavior in gas turbine inlet ducts possible. As soon as a droplet touches the air, and if the vapor pressure near the droplet surface is higher than the vapor pressure of the air far from the droplet (i.e., the droplet is in unsaturated air), evaporation of the droplet starts to occur. In order to balance the evaporation (mass transfer), the droplet has to lose a quantity of energy (heat transfer), which reduces the temperature of the droplet and of the air surrounding it. Depending on the droplet relative velocity, either natural or forced convection will occur. Natural convection occurs when the relative velocity of a droplet compared to the surrounding air is zero, and forced convection occurs when a relative velocity differential exists.

The effect of velocity on evaporation is not significant in our conditions (the effect is less than 10% for 50 microns droplet and further decreases when the droplet size diminishes) because the droplets are very small and their response time, in our experimen-

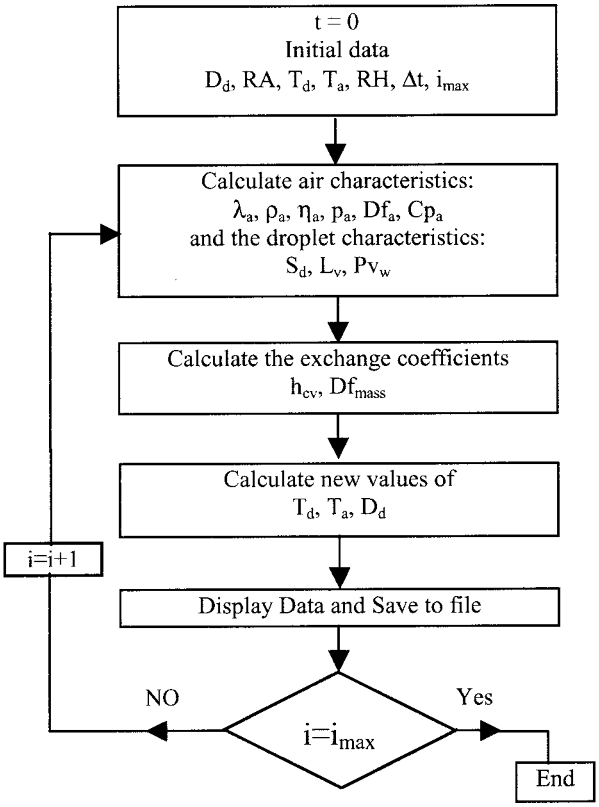


Fig. 14 Computational model for droplet evaporation

tal conditions, is lower than 10 ms [3]. This means that the use of the single droplet model gives a good estimation of the behavior of all the droplets in the duct.

6.1 Model Results and Discussion. The basic model was described in detail in Chaker et al. [1] and it has now been extended so it may be applied to the entire distribution of droplets. The droplet size distributions collected by the laser particle analyzer typically have 30 to 40 classes of droplet size. By calculating the evaporation time of each size class, it is possible to calculate the evaporation efficiency of the entire distribution. The basic procedure is as follows: First, we insert into the model the volume percentage of all the size classes and other initial data such as droplet temperature and the psychrometric parameters of the air. Second, we calculate the evaporation rate of each size class and then, by multiplying each class by the percentage of the volume of water evaporated, we calculate the new initial conditions and iteratively reinsert them in the model, until the final solution is derived.

The results are shown in Figs. 15, 16, and 17, which are discussed in detail below. Taking into account the different ambient psychrometric conditions that may typically exist with inlet air cooling in different climate zones, based on analysis done by Chaker et al. [15,16], three initial ambient conditions have been considered:

- 45°C (113°F) and 5% RH: Hot and dry (HD) desert weather conditions with evaporative cooling potential of 26°C (47°F);
- 30°C (86°F) and 55% RH: Typical temperate climate summer condition with cooling potential of 8°C (14.4°F);
- 15°C (59°F) and 80% RH: Cold and humid (CH) weather conditions with 2°C (3.6°F) of cooling potential.

6.1.1 Discussion of Evaporation Dynamics for Single Droplets. A set of curves that shows the behavior of individual fog droplets at the three different initial ambient conditions are given in Fig. 15 (note that Ta_i, and RH_i, refer to the initial air temperature and relative humidity, respectively). In the charts on the left side of the figure (*a*), (*c*), and (*e*), the abscissa is the initial droplet diameter (Dd_i) and the ordinate provides the time required to reach the humidity level given by the different curves.

In the charts shown on the right side (*b*), (*d*), and (*f*), the abscissa is again the initial droplet size and the ordinate is the droplet diameter at the point in time when the humidity given in the different curves is reached. The curves give an understanding of fog droplet evaporation as noted in the observations ahead:

- In examining Fig. 15(*a*), let us consider an initial droplet size of 20 microns. We enter the abscissa at 20, move upwards to an RH of 95% and find, on the ordinate, that reaching this RH took about 1.1 seconds. Moving to Fig. 15(*b*), and entering the abscissa, again at 20 microns, we move up to the 95% RH curve and, reading from the ordinate, we note that the droplet evaporated down to a diameter of about 4 microns.
- Doing the same exercise with Fig. 15(*c*) we see that with cooler and moister ambient conditions the time required to attain 95% RH with a 20-micron droplet increases to about 2 seconds. And, moving to Fig. 15(*d*) we see that evaporation to 95% RH reduced the 20-micron droplet to just 9 microns.
- In examining charts we see the critical importance of droplet diameter. For example, if we start with droplets of 40 microns, and assume that we have only 1 s residence time before the droplet reaches the compressor (typical for an inlet fogging system), we see that the relative humidity has been increased only to 60% and the remaining droplet is still more than 30 microns in diameter. Obviously an inlet fogging system that produced droplets in the 40-micron range would not be very efficient and would introduce large droplets to the compressor.
- We can see that ambient conditions play a major role in speed of evaporation. With typical summertime conditions for a temperate climate (Figs. 15(*c*) and (*d*)), the time required to reach 95% relative humidity increases to 2 s and the remaining droplet is 9 microns. Contrast these values with the values for the desert climate, given in the example above, where they were 1.1 s and 4 microns.

6.1.2 Discussion of Evaporation Efficiency for Entire Spray The volume frequency and cumulative volume for two initial ambient cases of 45°C (113°F) with 5% RH and 15°C (59°F) with 80% RH are shown in Figs. 16 and 17, respectively. These charts are based on a 1 s residence time, as that is typical for gas turbine inlet ducts. The charts were constructed using the curves of Fig. 15 and computing the evaporative efficiency for each different class of droplet sizes. This allows a determination of the evaporated water and nonevaporated water that would be expected from fog systems. Between 30 and 40 classes of droplet diameter were considered, depending on the range of droplet size distribution.

The most important quantitative factors for the fog spray are the surface area of water exposed for evaporation (which affects the evaporative cooling efficiency of the spray) and the size of the largest droplets (which affects the potential for compressor blade distress as well as giving an indication of the amount of water that will fall out in the duct) [2,3]. Given the above, the SMD, which is equal to the volume-to-surface-area ratio of the entire spray, and the Dv90 numbers are of primary importance. However, the presence of a large number of small droplets, which can represent a negligibly small quantity of injected water, can significantly reduce the D32 value. Since the total mass of small droplets is insignificant for the cooling process, using SMD alone can be misleading. When comparing two different nozzles it is best to use Dv90.

DV90 is relatively unaffected by numerically large populations of small droplets, which may represent a very small collective mass. Therefore, Dv90 was used to select typical droplet size distributions to characterize the evaporation efficiency of injected water droplets.

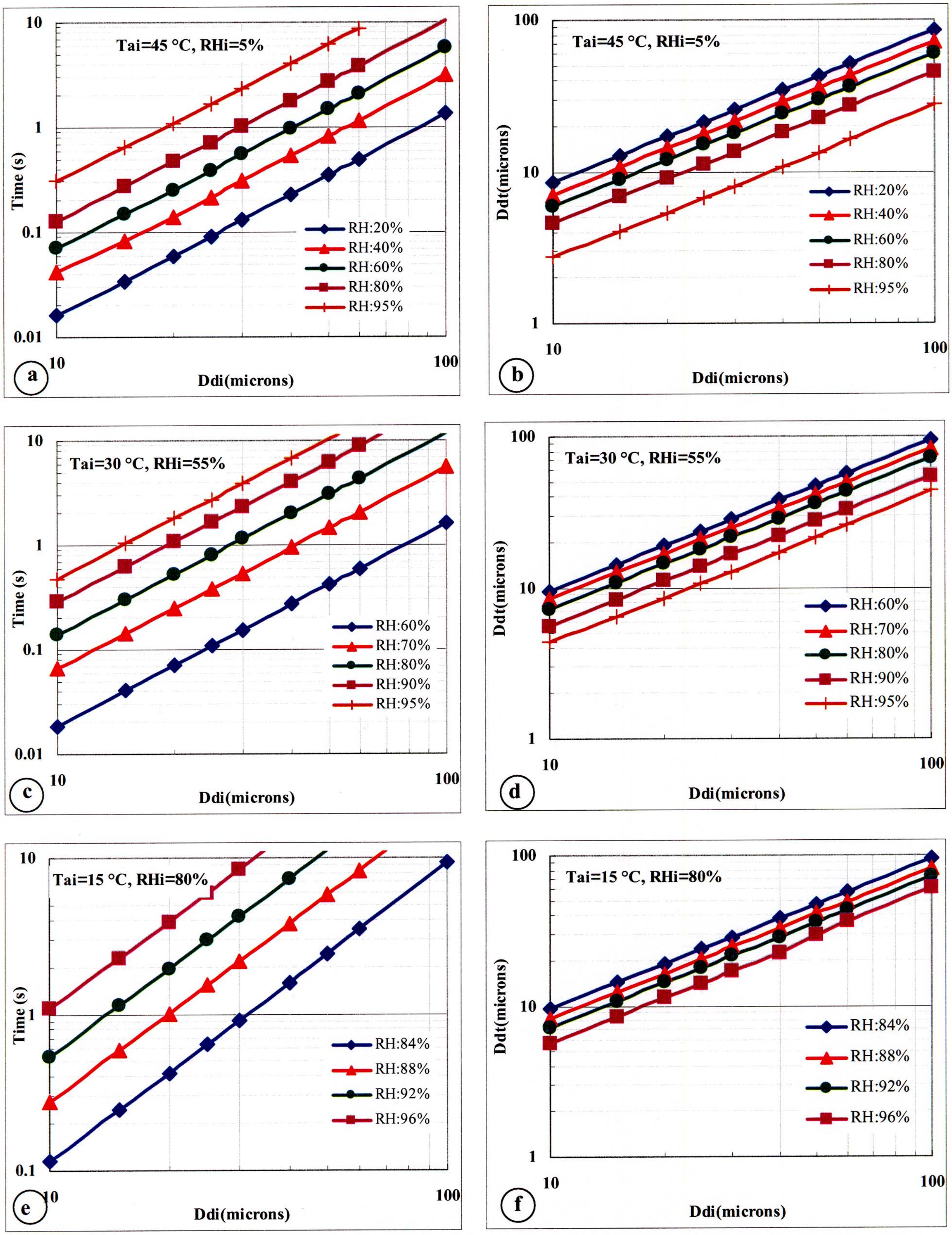


Fig. 15 Evaporation curves for single droplets showing residence time requirements, ending RH, and final droplet sizes for three ambient conditions

In Figs. 16 and 17, data have been provided for three different fog droplet distributions with Dv90 diameters of about 18, 28, and 46 microns. These distributions were chosen as being representative of nozzles that have been applied for inlet fogging (the 18-micron distribution being representative of that produced by Mee nozzles at 138 bar, or 2000 psi).

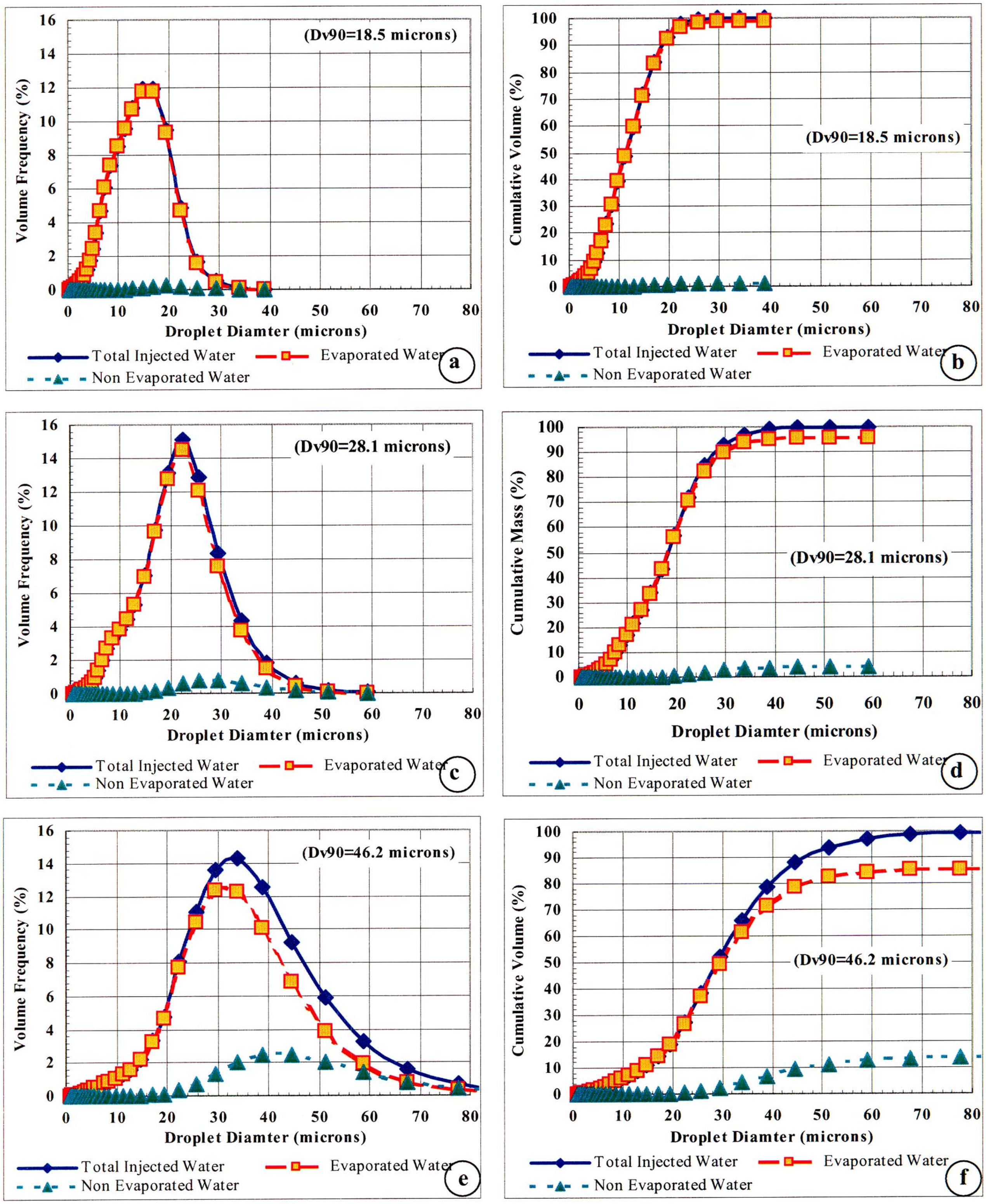


Fig. 16 Volume distributions showing total injected water, unevaporated water, and evaporated water for ambient conditions *of 45°C (113°F) with 5% RH and one second residence time.* The curves on the left (*a*, *c*, and *e*) show the volume frequency, and curves on the right (*b*, *d*, and *f*) provide cumulative volume.

In Figs. 16(*a*) and (*b*), the initial Dv90 diameter is 18.5 microns; it can be seen that the total injected water (the blue line) and the evaporated water (the red line) follow the same curve, and therefore the nonevaporated water (the green line) is at zero. In Figs. 16(*c*) and (*d*) with a Dv90 of 28.1 microns, we see that there is some separation of the blue and red lines. In Fig. 16(*d*), we see that the evaporated water curve never reaches more than 95%, so the unevaporated water is about 5%. The remaining spray has a Dv90 size of up to 39 microns. The reason for the increase in Dv90 (from 28 to 39 microns in this case) is the fact that the

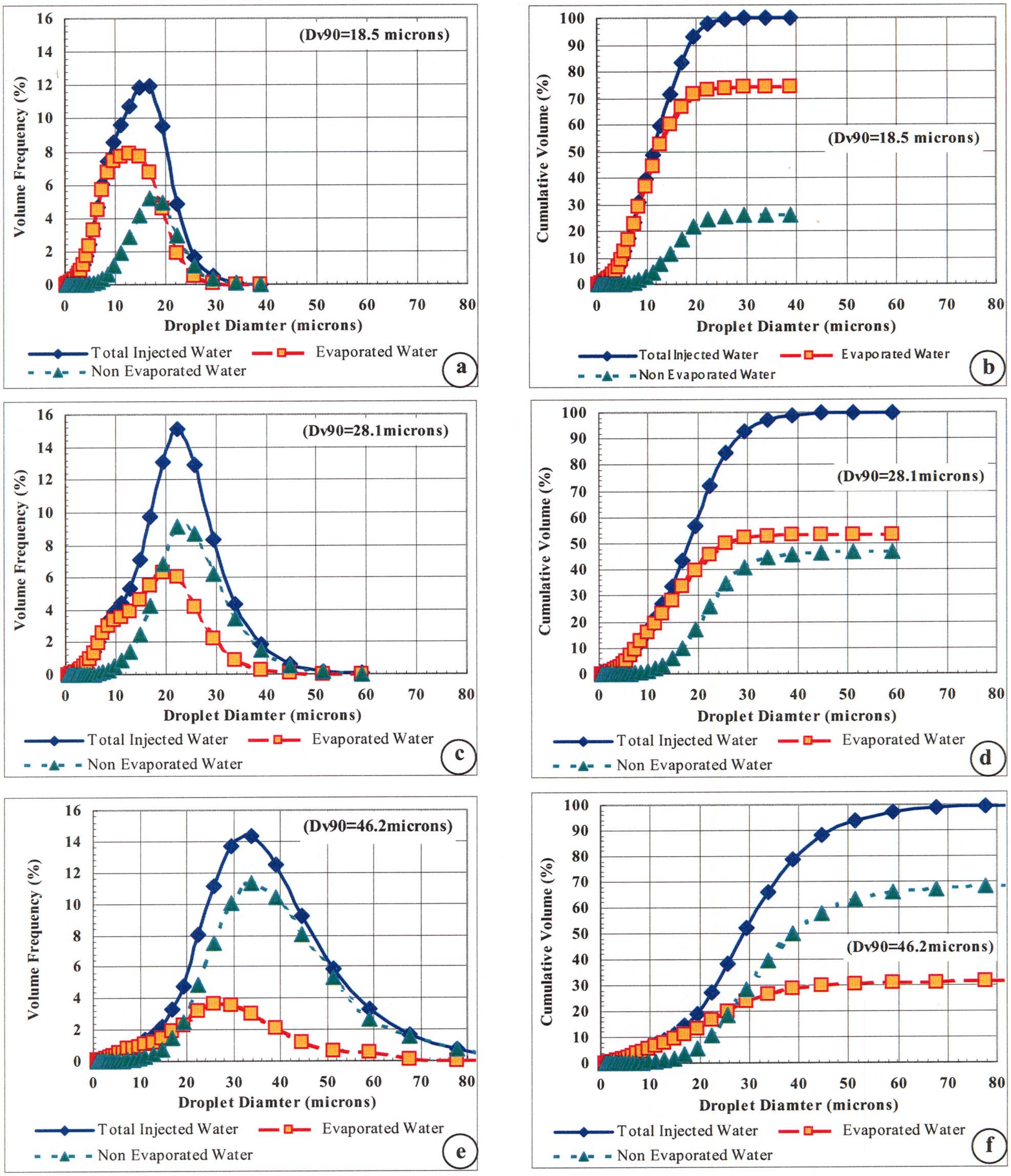


Fig. 17 Volume distributions showing injected water, unevaporated water, and evaporated water for ambient conditions of ***15°C (59°F) with 80% RH and one second residence time.*** The curves on the left (***a***, ***c***, and ***e***) show the volume frequency and curves on the right (***b***, ***d***, and ***f***) provide cumulative volume.

smaller droplets evaporate to zero, while the largest droplets get only slightly smaller. Figures 16(*e*) and (*f*) shows a Dv90 of 46.2 microns; the unevaporated water increases to 14% and the Dv90 size of the nonevaporated droplets is as high as 59 microns.

Figure 17 is similar to Fig. 16 except that the initial psychrometric conditions are much cooled and moister. The picture changes significantly. In examining Fig. 17(*a*), we see that the majority of the unevaporated water (the green line) consists of droplet sizes between 10 and 30 microns and the maximum volume percentage of unevaporated droplets is in the 17-micron size class. In looking at Fig. 17(*b*), one can see that 25% of the water did not evaporate. This is very interesting when viewed in comparison with the hot and dry climate conditions (Figs. 16(*a*) and (*b*)), where all the fog evaporated. Looking at the graphs of Figs. 17(*c*) and (*e*), where the initial droplet size is 46 microns, we see that the amount of nonevaporated water increases significantly (to almost 70%), as does the maximum final droplet size (80 microns).

It is important to note here that the size of unevaporated droplets for a Dv90 distribution of 18.1 microns varies between 10 and 30 microns, while this number increases to between 10 and 50 microns for a Dv90 distribution of 28.1 microns and reaches a value between 10 and 80 microns for a Dv90 distribution of 46.2 microns. In addition to the decrease in evaporation efficiency the considerably higher Dv90 may also have a negative impact on compressor blade life.

7 Considerations Relating to Positioning of Nozzle Manifolds

The position of the nozzle's manifold in the duct should be chosen with care. In general, there are three positions in the gas turbine duct where nozzle manifolds could be installed.

- Close to the inlet filter housing where the airflow velocity is around 2.5 m/s (500 fpm); this position is commonly used for evaporative fogging applications.
- After the silencer where the velocity is around 12.7 m/s (2500 fpm). This position is used for evaporative fogging and for combined overspray fogging.
- In the duct close to the axial compressor inlet, which is the typical overspray installation, where the residence time will be of the order of 0.2 s. The velocity here would also be close to 12.7 m/s (2500 fpm).

Installing the fog manifold close to the inlet filter housing, the airflow velocity is low, leads to longer residence time and, therefore, better evaporative cooling efficiency. However, the fog spray is poly-dispersed so the penetration velocity of the bigger droplets emitted from the nozzle orifice is higher than the penetration velocity of the smaller ones and, consequently, droplet collision and coalescence occurs.

By installing the nozzle manifold after the silencer, the coalescence effect is reduced significantly due to the fast response time of the smallest droplets to the high airflow velocity. Large and small droplets are separated into different flow paths and collisions are greatly reduced. The higher Weber number may also lead to secondary droplet breakup. A typical nozzle that produces a Dv90 of 25 microns at the lower velocity of the air-filter house will produce a Dv90 of 20 microns in the higher air velocity that exists after the silencer. By installing the nozzle's manifold after the silencer, however, the residence time of the droplets in the duct is typically reduced from 1 s to just 0.3 s.

Given the above, the position of the nozzle manifold in the duct should be chosen by taking into account the trade-off of droplet size and residence time in the duct. Since all the droplets are small enough to quickly take the velocity of the airflow, the effect of the velocity itself on the evaporation rate is negligible. In cases where maximum evaporation efficiency, minimum fallout, and no overspray are desired, locating the nozzles in the filter house will almost always be the better option.

An example is provided here for a hypothetical gas turbine with a mass flow rate of 500 Kg/s, with ambient psychrometric conditions of 30°C (86°F) and 20% RH and with a fog injection rate of 5 kg/s. Only 3 kg/s of water are required to saturate the 500 kg/s airflow, so the remaining 2 kg/s of water (0.4% of total airflow) becomes overspray.

By using the droplet size distributions measured at 2.5 m/s (filter-house velocity) and 12.7 m/s (duct velocity), calculations were made to determine the quantity of water that will be evaporated within 1.2 s (for the 2.5 m/s distribution) and 0.2 s (for the 12.7 m/s distribution).

The results for a residence time of 1.2 s are shown in Figs. 18 and 19. In looking at Fig. 18, the unevaporated water has droplet distribution sizes of 18–40 microns. The unevaporated water (overspray) will reach about 5% of the total injected fog, as can be seen in Fig. 19. The evaporation efficiency (using the active radius approach) after 1.2 s, with ambient conditions of 30°C (86°F) and 20% RH and using the size distribution measured at 2.5 m/s, is 94.4% (red curve in Fig. 19). The Dv90 as can be seen from Fig. 19 is 25 microns.

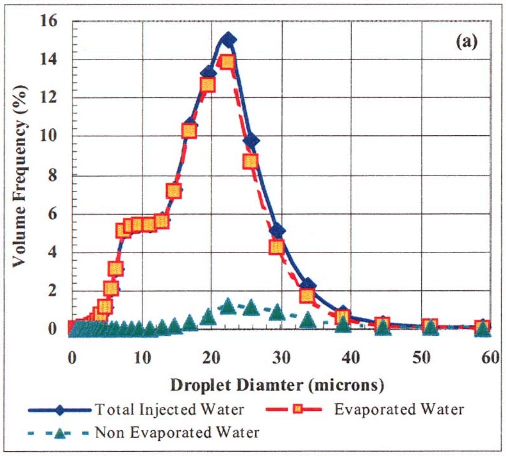


Fig. 18 Volume frequency of droplets from nozzles located in a low velocity region (filter house). Residence time is 1.2 s.

The corresponding figures for a shorter residence time of 0.2 s are shown in Figs. 20 and 21. The evaporation efficiency after 0.2 s, using the size distribution measured at 12.7 m/s, drops to 77.3% as shown in Fig. 21. The amount of unevaporated water is approximately 22% and the Dv90 is 20 microns.

In these conditions, the installation of the nozzle's manifold in the inlet filter housing leads to a better evaporation efficiency (94.4%) compared to the second position close to the axial compressor inlet with an evaporation efficiency of 77%.

By injecting 5 kg/s (0.4% overspray), it is clear that we can reach saturation in both cases (0.94*5=4.7 kg/s in the first case, and 0.77*5=3.9 kg/s in the second case) and only 3 kg/s is required to reach saturation. Therefore, the wet bulb temperature is reached in both cases.

7.1 Generalized Curve for Evaporative Efficiency.

In or-

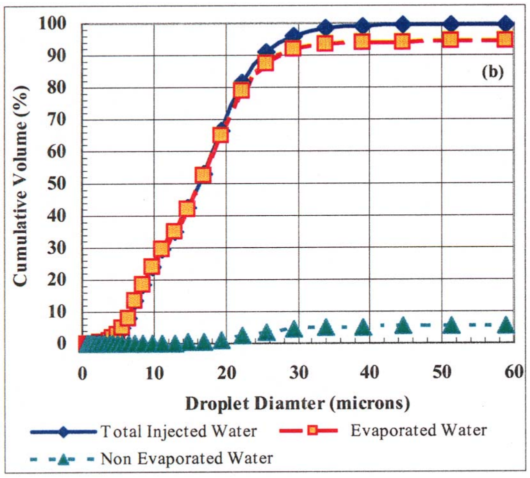


Fig. 19 Cumulative volume frequency of droplets from nozzles located in the low velocity region (filter house). Residence time is 1.2 s. Evaporation efficiency is 94.4%.

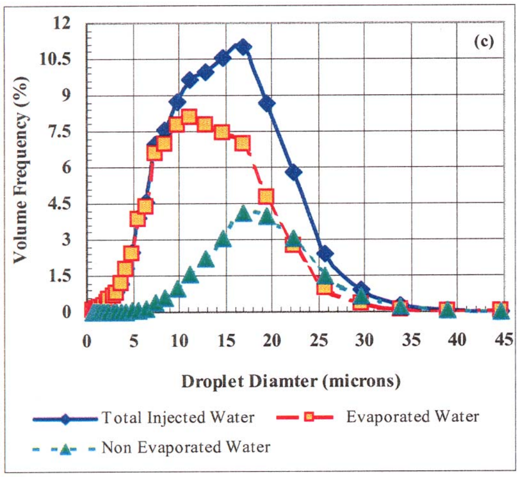


Fig. 20 Volume frequency of droplets for nozzle located in the high velocity region of the duct (after silencers). Residence time is 0.2 s.

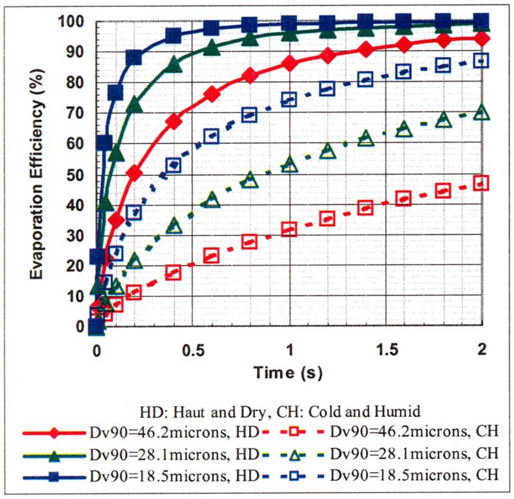


Fig. 22 Summary curve showing evaporative efficiency for a range of residence times and Dv90 droplet sizes

der to generalize the results showing evaporative efficiency and residence time, Fig. 22 has been constructed. Two sets of climate conditions have been considered defined as:

- Hot and dry (HD)—top three solid curves.
- Cold and humid (CH)—lower three dotted curves.

Considering a residence time of 1 s and assuming a hot and dry day, the evaporative efficiencies will range between 85% and about 100% for sprays with a Dv90 ranging from 18–46 microns. Even on a cold and humid day, this would hold true for Dv90 of 18.5 microns, while the evaporation efficiency decreases to around 32% for a Dv90 of 46.2 microns.

7.2 Experimental Verification of the Evaporative Efficiency Model. Experimental testing was conducted in the wind tunnel to validate the results of the theoretical evaporation efficiency model. The results of this are shown in Fig. 23. In this figure, the solid blue line is the evaporation efficiency predicted by the model and the experimental test points are also provided. It can be seen that there is a close correlation between the results predicted by the evaporation model [1] and the experimental results from experiments conducted in the wind tunnel at different airflow velocity and ambient conditions.

In examining the fit between the predictive model and the actual test results, it can be seen that the model slightly underpredicts evaporative efficiency at the lower residence time (for example the data point around 0.4 s residence time). The model slightly underpredicts efficiency for residence times ranging from 0.9–1.5 s. The reason for this is that the model assumes that each drop has the required volume to saturate the air around it but, in reality, this is not the case. Close to the nozzle the droplet population is very dense and also there will be some regions of untreated air between the nozzles. Therefore, the model tends to overestimate the evaporative efficiency for very short residence times. As the residence times increase, the small droplets tend to evaporate faster, thus increasing the relative humidity and making evaporation of the larger droplets more difficult. It is important to note that the deviations are exceedingly small; however, a physical explanation has been provided to understand the reasons for the deviation. Work is underway to extend the model further to incorporate these factors.

This model was verified using experimental data from a wind tunnel. The experiments were done without the presence of a silencer or any other obstructions such as trash screen or heating tubes. In real gas turbine ducts, the impact of large droplets with

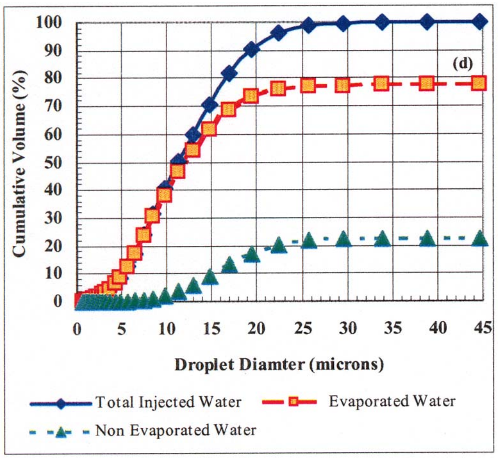


Fig. 21 Cumulative volume frequency of droplets from nozzles located in the high velocity region (after the silencers). Residence time is 0.2 s. Evaporation efficiency is 77%.

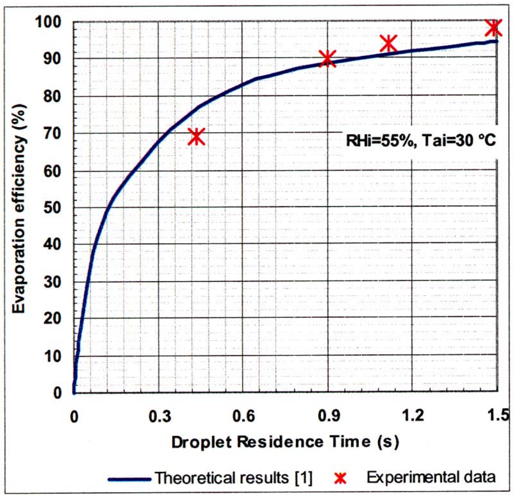


Fig. 23 Correlation between theoretical model prediction and experimental data

these obstructions may lead to a decrease in evaporation efficiency and an increase of the Dv90 of the nonevaporated fog droplets. Due to the large variations of geometry of gas turbine inlet configurations, it is not practical to simulate them all. Each configuration needs to be treated on a case-by-case basis. The model extension planned will incorporate these effects.

Depending on the inlet duct geometry, the airflow velocity may vary between the different regions in the inlet duct [17], and the fog nozzle should, therefore, be distributed in such a way to take into account this difference in airflow velocity. Such distributions will aid mixing and reduce temperature deformation at the compressor inlet.

8 Conclusions

This paper has provided the results of extensive experimental and theoretical studies conducted on impaction-pin fog nozzles used for gas turbine inlet air fogging and the dynamics of inlet fogging in general. It has been shown that inlet fogging nozzles produce smaller droplets at higher airflow velocities but that ambient humidity levels do not significantly affect droplet size, when measurements are taken within about 20 cm (7.9 in.) of the orifice. Droplet size tests were performed on nozzle with intentionally dislocated impaction pins, and it is shown that, unless gross impaction-pin dislocations are made, the droplet size remains essentially the same. The phenomenon of large-droplet generation (sprinkle) from both impaction-pin and swirl jet nozzles has been examined and in has been shown that, for the Mee impaction-pin nozzle, sprinkle goes away at the higher operating pressures and does not significantly impact fog system performance or operational safety. An improved numerical model for predicting fog evaporation rates is presented along with discussion and curves for different initial conditions. The curves make it possible to quantitatively analyze different fog sprays under various ambient climate conditions. The given data are applied to a discussion of the ideal location of fog nozzles in the gas turbine intake air duct and the effect that a particular location will have on the evaporative efficiency of the fog system, as well as the droplet size of fog that may enter the compressor. Experimental results for evaporation time were found to be compatible, within the range of experimental uncertainty, with the predictions obtained by the numerical model.

Nomenclature

Cp_a = specific heat of air ($J \cdot kg^{-1} \cdot °C^{-1}$)
Cp_d = specific heat of water ($J \cdot kg^{-1} \cdot °C^{-1}$)
D_{32} = Sauter mean diameter (microns)
D_d = droplet diameter (microns)
Df_a = mass coefficient of diffusion for air ($m^2 \cdot s^{-1}$)
Df_{mass} = coefficient of mass diffusion ($m \cdot s^{-1}$)
Dv_{90} = is a diameter for which 90% of the water volume in the spray is less than or equal to (microns)
h_{cv} = coefficient of thermal convective exchange ($W \cdot m^{-1} \cdot °C^{-1}$)
I = iteration in the evaporation dynamics model
Lv = latent heat of vaporization of water ($J \cdot Kg^{-1}$)
P_a = atmospheric pressure (Pa)
Pv_w = vapor pressure at the interface droplet-air (Pa)
RA = active radius (microns)
RH = relative humidity (%)
S_d = droplet surface (m^2)
T_a = temperature of air (°C)
T_d = droplet temperature (°C)
V_{rel} = droplet relative velocity ($m \cdot s^{-1}$)
We = Weber number

Greek

λ_a = thermal conductivity, air ($W \cdot m^{-1} \cdot °C^{-1}$)
μ_a = dynamic viscosity, air ($kg \cdot m^{-1} \cdot s^{-1}$)
ρ_a = density, air ($kg \cdot m^{-3}$)
γ_w = water surface tension ($N \cdot m^{-1}$)
η_{ev} = evaporation efficiency (%)
ΔT = temperature difference (°C)

Subscripts

a = air
d = droplet
t = time
i = initial

Appendix

Definitions of droplet diameters used in gas turbine inlet fogging applications and some parameters used in the paper:

- *Dv90* is a diameter for which 90% of the water volume in the spray is less than or equal to. A small value of this number indicates that a very small number of larger droplets is present. A small Dv90 minimizes the potential for impaction on obstructions and droplet fallout due to gravity (both of which reduce water pooling on the duct floor) and reduces the potential for compressor blade distress.
- *Sauter mean diameter (SMD or D32)* is the diameter of a hypothetical droplet whose ratio of volume to surface area is equal to that of the entire spray. Since it deals with surface area, Sauter mean diameter is a good way to describe a spray that is used for processes involving evaporation. To enhance droplet evaporation, one has to maximize the active surface area and minimize the internal volume of the droplet; thus, the lower the Sauter mean diameter, the more rapid the evaporation process.
- *Evaporation efficiency*. This is the percent of quantity of water evaporated compared to the total quantity of injected water.
- *Cumulative volume frequency of droplets*. This is the total volume of spray contained in droplets below a given diameter.
- *Active radius RA* is the volume around the droplet for which saturation conditions are calculated.

References

[1] Chaker, M., Meher-Homji, C. B., and Mee, T. R., III, 2002, "Inlet Fogging of Gas Turbine Engines-Part A: Fog Droplet Thermodynamics, Heat Transfer and Practical Considerations," ASME Paper No: 2002-GT-30562.
[2] Chaker, M., Meher-Homji, C. B., and Mee, T. R. III, 2002, "Inlet Fogging of Gas Turbine Engines-Part B: Fog Droplet Sizing Analysis, Nozzle Types, Measurement and Testing," ASME Paper No: 2002-GT-30563.
[3] Chaker, M., Meher-Homji, C. B., and Mee, T. R. III, 2002, "Inlet Fogging of Gas Turbine Engines-Part C: Fog Behavior in Inlet Ducts, CFD Analysis and Wind Tunnel Experiments," ASME Paper No: 2002-GT-30564.
[4] Meher-Homji, C. B., and Mee, T. R., 1999, "Gas Turbine Power Augmentation by Fogging of Inlet Air," Proceedings of the 28th Turbomachinery Symposium, Houston, TX, September 1999.
[5] Meher-Homji, C. B., and Mee, T. R., 2000, "Inlet Fogging of Gas Turbine Engines-Part A: Theory, Psychrometrics and Fog Generation and Part B: Practical Considerations, Control and O&M Aspects," ASME Paper Nos: 2000-GT-0307 and 2000-GT-0308.
[6] Kleinschmidt, R. V., 1946, "The Value of Wet Compression in Gas Turbine Cycles," Annual Meeting of the ASME, December 2–6, 1946.
[7] Wilcox, E. C., and Trout, A. M., 1951, "Analysis of Thrust Augmentation of Turbojet Engines by Water Injection at the Compressor Inlet Including Charts for Calculation Compression Processes With Water Injection," NACA Report No: 1006.
[8] Hill, P. G., 1963, "Aerodynamic and Thermodynamic Effects of Coolant Ingestion on Axial Flow Compressors," Aeronaut. Q., February, pp. 333–348.
[9] Arsen'ev, L. V., and Berkovich, A. L., 1996, "The Parameters of Gas Turbine Units With Water Injected Into the Compressor," Thermal Eng., **43**, No. 6, pp. 461–465.

[10] Nolan, J. P., and Twombly, V. J., 1990, ‘‘Gas Turbine Performance Improvement Direct Mixing Evaporative Cooling System,’’ ASME Paper No: 90-GT-368.

[11] Utamura, M., Kuwahara, T., Murata, H., and Horii, N., 1999, ‘‘Effects of Intensive Evaporative Cooling on Performance Characteristics of Land-Based Gas Turbine,’’ Proceedings of the ASME International Joint Power Generation Conference, 1999.

[12] Le Coz, J. F., 1998, ‘‘Comparison Of Different Drop Sizing Techniques On Direct Injection Gasoline Sprays,’’ 9th International Symposium On Application Of Laser Techniques To Fluid Mechanics, Lisbon, 13–16 July, 1998.

[13] Hinze, J. O., 1955, ‘‘Fundamentals of the Hydrodynamic Mechanism of Splitting in Dispersion Process,’’ AIChE J., **1**, No. 3, pp. 289–295.

[14] York, J. L., Stubbs, H. F., and Tek, M. R., 1953, ‘‘The Mechanism of Disintegration of Liquid Sheets,’’ Trans. ASME, **75**, pp. 1279–1286.

[15] Chaker, M., Meher-Homji, C. B., Mee, T., and Nicolson, A., 2001, ‘‘Inlet Fogging of Gas Turbine Engines—Detailed Climatic Analysis of Gas Turbine Evaporative Cooling Potential in the USA,’’ ASME Paper No. 2001-GT-526.

[16] Chaker, M., and Meher-Homji, C. B., 2002, ‘‘Inlet Fogging of Gas Turbine Engines—Detailed Climatic Analysis of Gas Turbine Evaporative Cooling Potential for International Locations,’’ ASME Paper No: 2002-GT-30559.

[17] Hoffmann, J., 2002, ‘‘Inlet Air Cooling Performance and Operation,’’ P.P. 222-227, T1-A-39, CEPSI 2002, Fukuoka, Japan.

Heimo Walter
e-mail: h.walter@tuwien.ac.at

Wladimir Linzer

Institute for Thermodynamics and Energy Conversion,
Vienna University of Technology,
Getreidemarkt 9,
1060 Vienna,
Austria

Flow Stability of Heat Recovery Steam Generators

This paper presents the results of theoretical flow stability analyses of two different types of natural circulation heat recovery steam generators (HRSG)—a two-drum steam generator—and a HRSG with a horizontal tube bank. The investigation shows the influence of the boiler geometry on the flow stability of the steam generators. For the two-drum boiler, the steady-state instability, namely, a reversed flow, is analyzed. Initial results of the investigation for the HRSG with a horizontal tube bank are also presented. In this case, the dynamic flow instability of density wave oscillations is analyzed.

[DOI: 10.1115/1.2179469]

Introduction

Steam generators with natural circulation have a wide range of applications, such as industrial heating processes or power cycles. Many of these natural circulation processes use the waste heat of gas turbines. In these so-called combined power cycles, the steam generator is arranged behind the gas turbine. Modern gas turbines for combined cycles are highly flexible in their mode of operation, i.e., concerning rates of startup, load change, and shutdown. Heat recovery steam generators (HRSG) arranged downstream of the gas turbine are forced to operate in such a way, that the gas turbine operation is not restricted by them. Therefore, they should be designed for a high cycling capability. Advantages of HRSGs with natural circulation evaporators are the reduced investment, maintenance, and operation costs due to the absence of a circulation pump.

HRSGs can be designed with vertical or horizontal tubes. The typical design of the circulation system of a natural circulation steam generator includes unequally heated tubes that are connected by headers with heated or unheated downcomers. Experience has shown that the most critical operational mode of a natural circulation HRSG is a hot startup or a heavy load change. In such cases, stagnation and/or reverse flow of the working fluid can occur. This type of flow condition is characterized as a static or steady-state instability. Criteria for the prediction of static instabilities were first presented by Ledinegg [1]. An overview and a classification of static and also dynamic instabilities in natural and forced circulation systems were presented by Bouré et al. [2].

A variety of computer codes have been written that deal with the evaluation of the static behavior of steam generators. As such, this performance characteristic has become a standard procedure in the design of many new natural circulation boilers. As a result of such calculations, the mass flow distribution of the working medium in the tube network, as well as other detailed information, such as density, pressure, and temperature, are available. The first part of the paper presents the results of a steady-state stability analysis based on a simple model of a natural circulation steam generator.

Figure 1 depicts a model of a natural circulation system. The boiler consists of a drum, an unheated downcomer, a lower header and two riser systems with unequally heated sections. The tubes of the two riser systems are all connected to the drum. The drum pressure of the boiler is 80 bars. It can be shown that for such a system, the number of stable solutions depends on the number of branches of the tube network.

Figure 2 shows the results of a steady state stability analysis for the circulation system shown in Fig. 1. The vertical axis of Fig. 2 represents the mass flow in heated riser system 1 $\dot{m}_1$ (i.e., the system that is exposed to a lower heat flux as riser system 2); the horizontal coordinates represent the heat flux for the two different heated riser systems.

In Fig. 2, the gray-shaded surface defines the area of possible mass flow rates $\dot{m}_1$ of the working fluid dependent on the heat flux to that is applied to both riser systems. For example, it can be seen that for a constant heat flux for the higher heated riser system 2 of $\dot{q}_2 = 320\ \mathrm{kW/m^2}$ and a decreasing value of heat flux for the lower heated riser system 1, only one solution (e.g., $V=2$, point B) with upward flow in both riser systems exists if the heat absorption ratio V is smaller than 7 (point A).

The heat absorption ratio V is defined as

$$V = \frac{\dot{q}_2}{\dot{q}_1} \tag{1}$$

With increasing heat absorption ratios (values higher than 7), such operational conditions for the boiler lead to additional solutions for the circulation distribution of the working medium. For such conditions, three solutions for the mass flow distribution can be found (see Fig. 2, $V=10$, points C, D, E). Two of them are stable (points C and E), while the third solution is unstable (point D). The two stable solutions are upward and downward (reverse) flow of the working fluid in the lower heated riser system 1.

The shape of the water circulation surface depends on the geometry of the boiler bank and the overall heating conditions imposed on the boiler. Accordingly, each situation and boiler configuration must be analyzed separately. Static analysis methodology can only determine if flow reversal is possible under steady state conditions. But this procedure is not able to predict whether flow reversal actually occurs.

Additional information regarding the results of a stability analysis to prevent reverse flow under hot startup conditions for the natural circulation system shown in Fig. 1 are presented in a separate paper (see [3]).

Density wave oscillations (DWO) are the most frequently observed type of the dynamic instabilities. The DWO is characterized by large amplitudes and a nearly sinusoidal period. A density wave oscillation is a low-frequency oscillation of the mass flow where the period is approximately between one and two multiples of the time required for a fluid particle to pass through the tube or channel. The oscillation of the pressure and the mass flux are in phase. For further information about this type of instability, see Ref. [2] or [4].

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received October 1, 2003; final manuscript received March 1, 2004. Review conducted by A. J. Strazisar. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Vienna, Austria, June 13–17, 2004, Paper No. GT2004-53040.

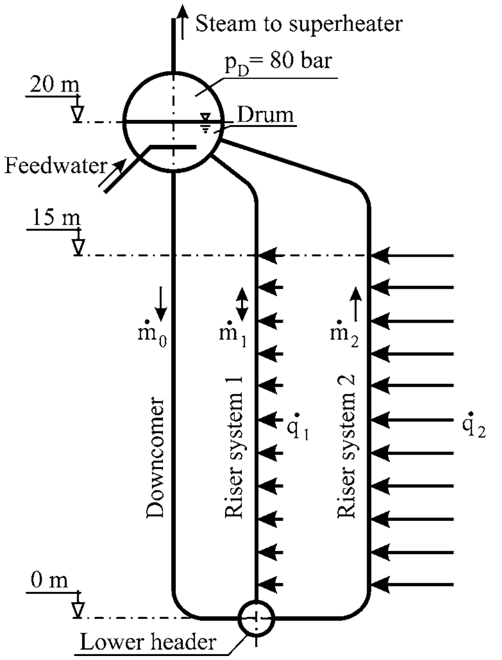


Fig. 1 Model of a natural circulation boiler

The following sections of this paper present the results of theoretical dynamic stability analysis for different HRSG designs. These studies incorporate the influence of the boiler geometry. These investigations were carried out for a two-drum HRSG and a HRSG with a horizontal tube bank at low pressure. Both steam generators are natural circulation boilers. The mathematical boiler model described below was used for all investigations.

Mathematical Modeling and Numerical Method

The computer program DBS[1] used in these evaluations was designed to analyze the dynamic behavior of steam generators, especially natural circulation HRSGs. In the following, a short description of the mathematical model for the tubes and the header (the so-called tube-header model) will be given.

Model of Fluid Flow in the Tube. The mass flow in the tubes of a steam generator can be assumed to be one-dimensional, as the length of the tubes is much greater compared to their diameter. For the model under consideration, the topology of the tube network, the number of parallel tubes, the geometry in terms of outer diameter and wall thickness of the tubes, the fin geometry, and the dimensions of the gas ducts are necessary. Furthermore, the thermodynamic data, such as mass flow, pressure, and temperature, of the heat exchanging streams are used as input data.

The mathematical model [5] for the working medium is one-dimensional in flow direction, uses a homogeneous equilibrium model for two-phase flow, and applies a correction factor for the two-phase pressure loss according to Friedel [6]. For a straight tube with constant cross section, the governing equations in flow direction are written for the mass and momentum conservation of the fluid

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho w}{\partial x} = 0 \qquad (2)$$

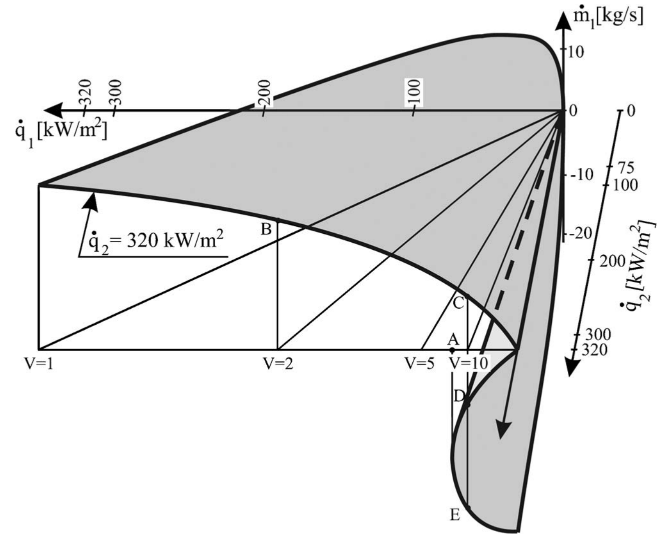


Fig. 2 Mass flow in the lower heated riser system 1 for different heat absorption ratios *V* as three-dimensional surface [3]

$$\frac{\partial \rho w}{\partial t} + \frac{\partial \rho w w}{\partial x} = -\frac{\partial p}{\partial x} - \rho g_x + \left(\frac{\partial p}{\partial x}\right)_{\text{friction}} \qquad (3)$$

The density ρ and the velocity w are averaged values over the cross section of the tube.

Considering the fluid flow in steam boilers, the thermal energy is much higher than the kinetic and the potential energy as well as the expansion work. Therefore, the balance equation for the thermal energy can be simplified to

$$\frac{\partial \rho h}{\partial t} + \frac{\partial \rho h w}{\partial x} = \dot{q}\frac{U}{A} \qquad (4)$$

The heat exchange between the fluid and the tube wall is governed by Newton's law of cooling, and the heat transfer through the wall is assumed to be in the radial direction only. The heat transfer models used in DBS for the single and two-phase flow of the working medium includes correlations for horizontal as well vertical tubes and is described, in detail, in [5].

Model of Header. For the calculations, the following assumptions may be made for the header:

- Assuming that the distribution of the thermodynamic state of the header is homogeneous, the collector can be seen as one control volume for the calculation.
- The gravity distribution of density and pressure of the fluid inside the header can be neglected because the vertical dimension of the header is small compared to that of the remaining tube system.
- The huge difference in the cross-sectional area of the header and the connected tubes causes strong turbulence, avoiding, thus, a segregation of the fluid in the header.

Because the headers are assumed to be a single control volume, the equations for the mass and energy balance are ordinary differential equations with time t as independent variable

$$\frac{d}{dt}(\rho_C V_C) = \sum_j \rho_j w_j A_j - \sum_k \rho_k w_k A_k \qquad (5)$$

$$\frac{d}{dt}(\rho_C h_C V_C) = \sum_j \rho_j w_j h_j A_j - \sum_k \rho_k w_k h_k A_k \qquad (6)$$

The variables of the header are denoted with the index C; j represents values at the header entrance and k values at the outlet. V_C

---

[1] The program "Dynamic Boiler Simulation" (DBS) was developed at the Vienna University of Technology, Institute for Thermodynamics and Energy Conversion.

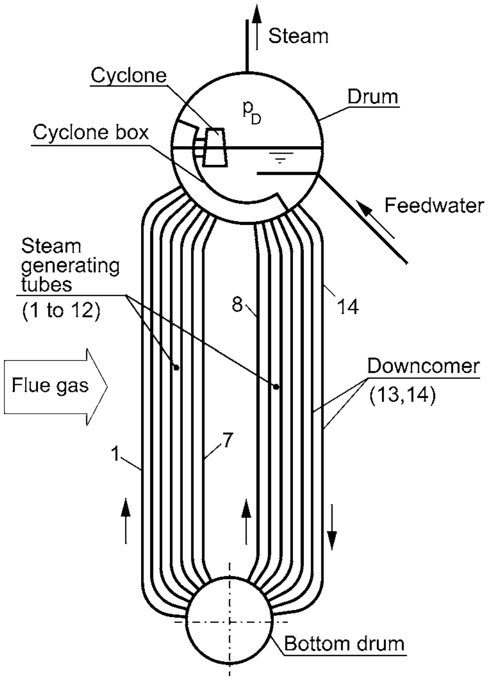


Fig. 3 Model of the two-drum boiler

is the volume of the fluid inside the header and A the cross-sectional area of the connected tubes. Similar to the treatment of the fluid flow in the tube, kinetic energy as well as expansion work is neglected in the energy balance.

Because momentum is a vector quantity having the direction of the tube axis, the momentum fluxes must not be added arithmetically but rather as vectors. This is the case at the inlet and outlet of the headers, where the individual tubes are connected under different angles. The velocity of the fluid in the header is rather small compared to that inside the tubes. Thus, it can be assumed that the momentum of the fluid will be lost at the inlet of the header and has to be rebuilt at the outlet. Based on this assumption, the momentum balance of the header reduces to a pressure balance. The changes of the momentum at the inlet and outlet can be taken into account by a pressure loss coefficient ζ

$$p_C = p_j - \frac{\zeta_j}{2}\rho_j w_j |w_j| \tag{7}$$

$$p_C = p_k + \frac{\zeta_k}{2}\rho_k w_k |w_k| \tag{8}$$

The discretization of the partial differential equations for the conservation laws was done with the aid of the finite-volume method. The pressure-velocity coupling and overall solution procedure are based on the SIMPLER algorithm [7]. To prevent checkerboard pressure fields a staggered grid is employed and for the convective term the UPWIND scheme is used.

Model of the Flue Gas. For the description of the flue gas, the one-dimensional partial differential equation of the conservation law for the energy is used. The momentum balance for the flue gas is neglected. The flue gas mass flows are calculated quasi-stationary, while the energy balance is calculated unsteady. The discretization of the energy balance is done corresponding to the finite-volume method. The convective heat transfer coefficient between the flue gas and the tubes can be calculated with different correlations for plain or finned inline or staggered tube banks (see [5]).

Two-Drum Natural Circulation Boiler. Figure 3 shows the model of a natural circulation two-drum boiler, which is arranged behind a gas turbine. The evaporator of the boiler consists of two downcomers (tubes 13 and 14), a bottom drum, the bundled heating surface with 12 parallel tube paths, a cyclone box, a cyclone separator, and an upper drum. The downcomer and the riser tubes are finned tubes. The difference in height between the centerlines of the drums is 6.5 m.

Saturated water (no subcooling) leaves the drum through the downcomers and enters the riser tubes through the bottom drum. The water-steam mixture leaving the riser tubes is collected in the cyclone box and passes through the cyclone where steam and water are separated. In comparison to downcomer tubes 13 and 14 the cyclone separator represents an additional flow resistance.

The flue gas enters the steam generator at the first riser tube (tube 1) and leaves the bundle heating surface at the second downcomer (tube 14). For the stability analysis of the two-drum boiler, the following three different geometric configurations have been analyzed:

- No cyclones and cyclone box are included in the drum and all evaporator tubes are finned tubes.
- No cyclones and cyclone box are included in the drum, and all riser tubes (tubes 1–12) are finned tubes while both downcomers are plain tubes.
- Cyclones and a cyclone box are included in the drum and all evaporator tubes are finned tubes.

In addition to this set of test cases, a second set of calculations has been performed. In this second set, the difference in height between the centerlines of the top and bottom drum is changed from 6.5 m to 11.6 m. This doubles the heated length of the tubes in the tube bank. To achieve the same thermal conditions, the mass flow of the flue gas is also doubled. For these conditions the steam production is approximately doubled.

At steady state, the circulation ratio for the case without cyclones was ~18 for a heated tube length of 5.1 m. With the arrangement of cyclones, and accounting for their pressure drop, the circulation ratio was reduced to ~14. For the cases with doubled heated tube length, the circulation ratio was reduced to ~6.5.

Initial and Boundary Conditions

For all test cases, the following initial conditions for the dynamic simulation of the hot startup of the two-drum boiler are used:

- The steam generator is filled with water near boiling condition.
- The pressure distribution of the working medium in the tube network of the boiler is affected by gravity.
- The velocity of the fluid at the start of the calculation process is equal to zero.
- The initial fluid temperature in the evaporator of the boiler is identical to the boiling temperature at drum pressure.
- The initial drum pressure of the steam generator is 12 bars.
- The drum pressure is constant during the whole simulation.
- The flue gas temperature and mass flow are given as a function of time (see Fig. 4), and are input boundary conditions for the simulation.
- The water level is controlled and held at the centerline of the drum.

Figure 4 shows the time-dependent flue gas temperature (full line) and flue gas mass flow (broken line) for the hot startup of the two-drum steam generator. The flue gas reaches full load ~320 s after the start of the simulation and remains constant at full load. The total time for the simulation is 1000 s. The flue gas temperature and mass flow ramps are the same for all simulations.

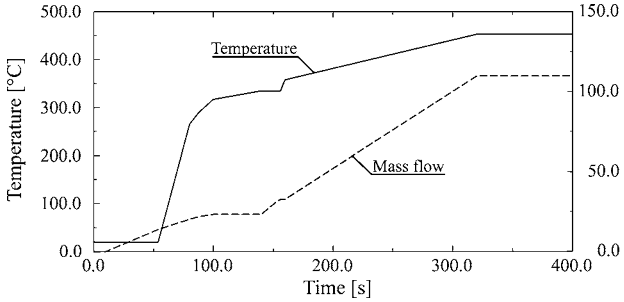


Fig. 4 Temperature and mass flow of the flue gas

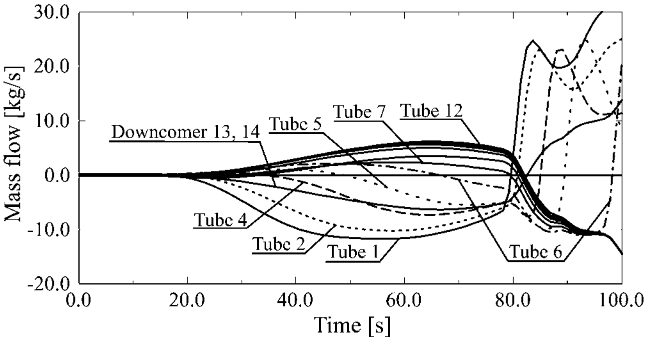


Fig. 6 Mass flow in various tubes of the two-drum boiler: detail of the first 100 s of the results shown in Fig. 5

Results of Dynamic Simulation for Two-Drum Boiler

The next several sections present the results of the two-drum boiler simulation. The following convention is used for the sign of mass flow in all figures:

- A positive sign in the riser tubes describes a flow direction from the bottom drum to the upper drum.
- A positive sign in the downcomer tubes indicates a flow direction from the upper drum to the bottom drum.

The first set of results is for the boiler with a difference in height of 6.5 m between the drum centerlines.

Results of Simulation With a Difference Height of 6.5 m

Model With Finned Downcomer Tubes and Without Cyclones. Figure 5 shows (as a result of the dynamic simulation for the hot startup) the mass flow distribution of the working fluid at selected points. Since the boiler is under the pressure of 12 bars, in the first phase of the hot startup the temperature of the flue gas (30°C, see Fig. 4) is lower than that of the tube material. Therefore, the flue gas is heated up and the tubes are cooled down. During this initial period, there is low circulation in the evaporator tubes. The working fluid in the more intensively cooled evaporator tubes (tubes 1–6) flows downward, while the mass flow in evaporator tubes 7–12 and both downcomer tubes is directed upward (Fig. 6). The direction of heat flow (between gas and tubes) changes with increasing mass flow and temperature of the flue gas.

With the beginning of the steam production in the most heated evaporator tube 1 (~80 s after the start of the simulation), the direction of mass flow of the working medium in this tube changes from downward to upward. Approximately 100 s after startup, the direction of mass flow in the first six tubes of the evaporator is upward. The direction of mass flow in evaporator tubes 7–12 as well as in both downcomer tubes (tubes 13 and 14) is directed downward.

Between ~100 s and 170 s, the heat flux from the flue gas to the evaporator tubes 7–12 is small and, consequently, the mass flow in these tubes remains in the downward direction. At about 170 s after the start the heat flux to tube 7 is high enough to start the steam production. With the beginning of steam production in tube 7, the density of the working medium decreases and the lifting or buoyant force increases. This results in a rapid decrease in mass flow in tube 7. With a further increase of the heat flux to tube 7, the mass flow of the water-steam mixture changes flow direction from downward to upward. As can be seen in Fig. 5, the same process occurs for evaporator tubes 8–11 in the time period between 170 s and 300 s.

For tube 12, the heat flux is not high enough to change the flow direction and, therefore, the tube operates as a third downcomer (even at steady state).

Figure 7 shows the time evolution of the pressure difference (due to height Δp_H) for tubes 1, 11, and 12, as well as the total pressure difference between the bottom and the upper drum Δp_{TD}. The distance between the curves Δp_{TD} and Δp_H in Fig. 7 represents the sum of the pressure differences due to friction and acceleration.

It can be seen that at the beginning of the hot startup, the pressure difference, due to height (static head) of tube 11, is higher than the total pressure difference between the bottom and the upper drum and, therefore, the flow is directed downward. With increasing steam production in tube 11, the static head of this tube decreases. Between 280 s and 300 s after start, the increase in steam production in tube 11 reduces the static head and the line of Δp_H crosses the trace of Δp_{TD}. During this period, the mass flow of the working medium in tube 11 changes flow direction from downward to upward.

The situation for tube 12 is different from that of tube 11. At the beginning of the simulation, the total pressure difference between the two drums is smaller than the pressure difference, due to

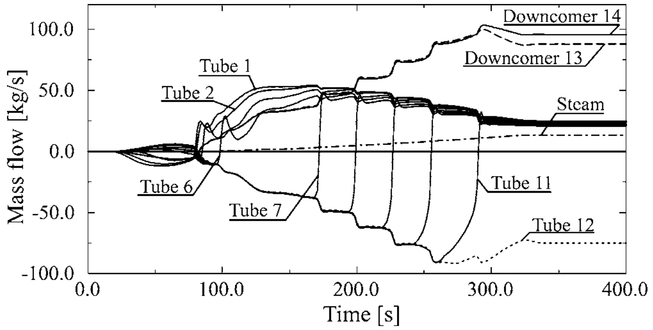


Fig. 5 Mass flow in various tubes of the two-drum boiler: dynamic behavior without cyclones and finned downcomer tubes (case 1)

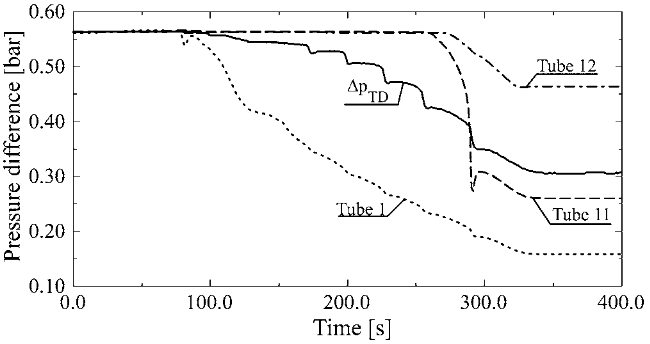


Fig. 7 Pressure difference in various tubes of the two-drum boiler: dynamic behavior without cyclones and finned downcomer tubes (case 1)

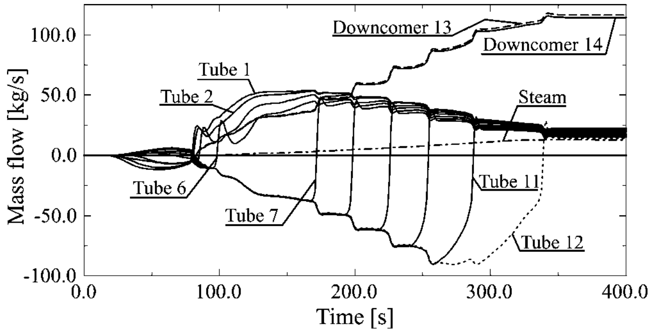


Fig. 8 Mass flow in various tubes of the two-drum boiler: dynamic behavior without cyclones and plain downcomer tubes (case 2)

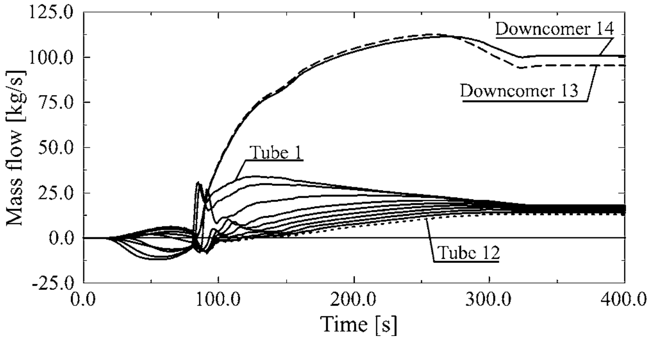


Fig. 10 Mass flow in various tubes of the two-drum boiler: dynamic behavior with cyclones and finned downcomer tubes (case 3)

height, of tube 12. With the increase in steam production in tube 12, the static head also starts to decrease, but the steam production is insufficient and, therefore, the mass flow of the water-steam mixture remains in the downward flow direction.

Model With Plain Downcomer Tubes and Without Cyclones. The calculated mass flow distribution at selected points of the two-drum steam generator without cyclone separators but with plain downcomer tubes (test case 2) is shown in Fig. 8. Compared to test case 1 (model with finned downcomer tubes), Fig. 8 clearly demonstrates the influence of the unfinned downcomer tubes on the circulation behavior.

The time evolution of the mass flow in the different tubes of the boiler is, in the first period of the hot startup (up to ~290 s), similar to that shown in Figs. 5 and 6. The flow in the first 11 riser tubes is directed upward after ~290 s, while flow in tube 12 is directed downward. In the period after 290 s (the start of the steam production in tube 12), the mass flow in the downcomer in test case 2 increased while the mass flow in test case 1 decreased (see Fig. 5). In test case 2, this results in a smaller mass flow rate and a higher steam production in tube 12. This higher steam production leads to an increase of the lifting force, which reduces the velocity of the water-steam mixture. This, in turn, results in a flow stagnation followed by a change of flow direction in tube 12 from downward to upward.

Figure 9 shows the static head Δp_H for tubes 1, 11, and 12 of the boiler as well as the total pressure difference between the bottom and the upper drum for test case 2. The influence of the plain downcomer tubes on the development of Δp_{TD} is evident. The time evolution of the pressure differences are similar to that shown in Fig. 7. Only the situation for tube 12 is different. In test case 1, Δp_H does not cross the curve of Δp_{TD}, whereas in test case 2, the steam production is high enough so that the curve for the static head of tube 12 crosses the curve for the total pressure difference between the bottom and the upper drum. In this time period, the mass flow in tube 12 changes flow direction from downward to upward.

Model With Cyclones and Finned Downcomer Tubes. Figure 10 shows the calculated mass flow distribution for the two-drum boiler model with cyclones and finned downcomer tubes (test case 3). During approximately the first 80 s after starting the simulation, the development of the mass flow in the evaporator tubes of the boiler is similar to the test cases 1 and 2. After 80 s, the mass flow distribution in the riser tubes for test case 3 is entirely different in than the other two cases. After ~150 s the flow in all the riser tubes is directed in the upward sense.

It should be noted that the two downcomer tubes are placed outside of the cyclone box while the riser tubes enter the cyclone box (see Fig. 3). Compared to test cases 1 and 2 (models without cyclones), Fig. 10 clearly demonstrates the influence of the cyclones on the circulation behavior. In the case of a reverse flow in a riser tube for test cases 1 and 2, saturated water leaves the drum, whereas in test case 3, a water-steam mixture leaves the cyclone box and enters the riser tube. Because of the lower fluid density (which enters the tube) in case 3, the lifting force increases and the fluid velocity decreases. When compared to the other two test cases, this results in a shorter time period of reverse flow and in a smaller mass flow rate in the riser tubes, which have lower heat flux during the period of reverse flow.

Figure 11, for test case 3, shows the static head Δp_H for tubes 1, 12, and both downcomer tubes 13 and 14, as well as the total pressure difference between the upper and the bottom drum. The influence of the cyclones on the development of Δp_H and Δp_{TD} is clearly seen. The time evolution of Δp_H of tube 12 during the first period of the hot startup is higher but very close to the curve of

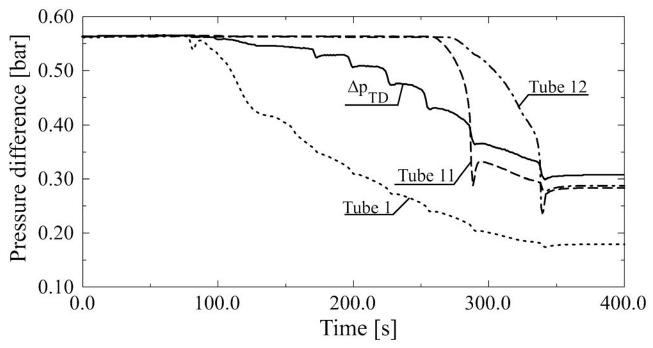


Fig. 9 Pressure difference in various tubes of the two-drum boiler: dynamic behavior without cyclones and plain downcomer tubes (case 2)

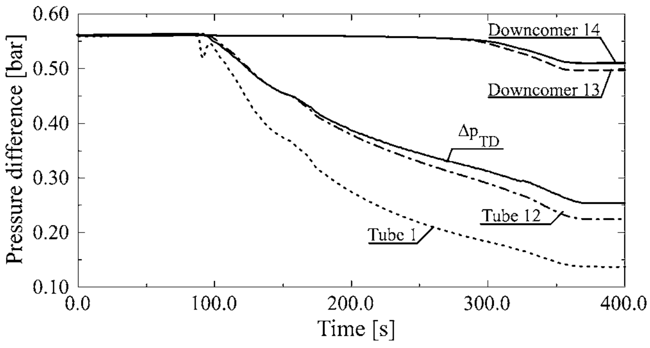


Fig. 11 Pressure difference in various tubes of the two-drum boiler: dynamic behavior with cyclones and finned downcomer tubes (case 3)

Δp_{TD}. Because of the small mass flow rate, the density of the working medium decreases and, therefore, the static head also decreases. Between 140 s and 150 s after the start of the simulation, the static head of tube 12 crosses the curve of the total pressure difference between the upper and the bottom drum. In this period, the mass flow of the working medium in tube 12 changes the direction from downward to upward.

Results of Dynamic Simulation With a Changed Boiler Height to 11.6 m. The results of the dynamic simulation of the two-drum boiler model without cyclone and plain downcomer tubes and the model with cyclones included and a height of 11.6 m between drum centerlines are similar to those presented in test cases 2 and 3. The time-dependent development of the mass flow, in the tubes of the steam generator, is the same as for test cases 2 and 3. Because of the higher mass flow of flue gas, the mass flow of the water-steam mixture in the individual tubes, as well as the steam quality at the connection point to the drum, increases.

The results for the mass flow distribution obtained with the model without cyclones and finned downcomer tubes differs from the calculation results with a drum distance of 6.5 m (test case 1). The steam quality in tube 12 is significantly higher for the case of a drum centerline distance of 11.6 m. Therefore, this tube also operates as riser contrary to the case with a 6.5 m drum centerline distance.

Summarizing the results of the simulation for the two-drum type boiler it can be stated as follows:

- Since flow reversal in any of the tubes should be avoided, the implementation of cyclones or similar separation devices is the most favorable design (test case 3).
- The reduction of the heat absorption for the downcomer tubes (last two rows in the bundle) has a stabilizing effect. With this measure, reverse flow in the lowest heated riser tube (tube 12) can be avoided at high loads.

HRSG With a Horizontal Tube Bank

The modern design of natural circulation HRSGs allows compact boiler designs. Horizontal tube banks are built with tube lengths up to 20 m. For high efficiency combined-cycle HRSGs, three pressure stages are frequently employed. The low-pressure stage in combination with the long horizontal tubes (which have no preferred flow direction for the working fluid) leads to more unstable operation conditions than HRSG designs with vertical tubes. Static instabilities, such as reverse flow and/or flow stagnation, as well as the dynamic instabilities, such as pressure drop instabilities or density wave oscillations, can occur (for further information, see Ref. [2]).

The model for the HRSG with horizontal tubes is shown in Fig. 12. The steam generator consists of a downcomer with a siphon at the lower end, a lower header, a bundled heating surface with four parallel tube paths, an upper header, a riser, and a drum. The heated length per layer of the horizontal tubes is 20 m, the difference in height between the centerline of the drum and the lowest point of the siphon is 10.5 m. The difference in height between the two headers as well as between the lowest point of the siphon and the lower header is 1.4 m. The flue gas enters the bundle heating surface at the bottom.

For the stability analyses of this HRSG the following different geometry configurations have been analyzed:

- Test case 4—All tubes of the bundled heating surface are finned tubes with an outer diameter of d_a=48.3 mm. The number of fins per meter is 236, the fin height is 12.7 mm, and the distance between the fins is 4.5 mm.
- Test case 5—The outer tube diameter of the first four layers of the bundled heating surface was changed from d_a =48.3 mm to d_a=44.5 mm (including the connecting tubes between the lower header and the finned layer tubes), while the other tube dimensions are the same as for test case 4.
- Test case 6—The outer tube diameter of the connecting tubes between the lower header and the finned layer tubes of layers 1–4 was changed from d_a=48.3 mm to d_a =44.5 mm, while the other tube dimensions are the same as for test case 4.
- Test case 7—The outer tube diameter of the connecting tubes between the lower header and the finned layer tubes of layers 1–4 was changed from d_a=48.3 mm to d_a =44.5 mm, while the other tube dimensions are the same as in case 4. Additionally, an orifice is installed at the inlet of all changed tubes.

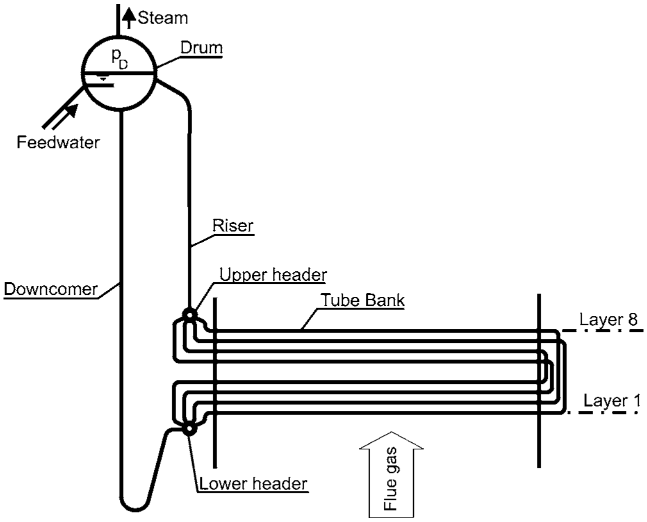


Fig. 12 HRSG with a horizontal tube bank

The circulation ratios for the different test cases of the HRSG with horizontal tubes is between 4.03 (test case 5) and 4.58 (test case 6).

Initial and Boundary Conditions. For the dynamic simulation of the hot startup, the same initial conditions are used as for the two-drum steam generator. The initial conditions are described above, in detail, and will not be repeated in this section.

As boundary condition for all boiler configurations the drum pressure is constant during the whole simulation and equal to 12 bars. The time evolution of the flue gas temperature is identical to that shown in Fig. 4. The values of the flue gas mass flow are adjusted to the boiler configuration and must be doubled. The water level of the drum is also controlled at the drum centerline.

Results of Dynamic Simulation for the HRSG With Horizontal Tube Bank

These next sections present the results of a theoretical stability analyses. For the mass flow direction, the following conventions are used in all figures:

- A positive sign for the mass flow in the evaporator tubes describes a flow direction from the lower header through the tube bank to the drum.
- A positive sign for the mass flow in the downcomer tubes describes a flow direction from the drum through the siphon to the lower header.

Results of Simulation With an Outer Tube Diameter of d_a =48.3 mm at All Layers. Figure 13 shows the mass flow distribution at selected points of the HRSG with a horizontal tube bank. The outer tube diameter for the simulation of the hot startup was 48.3 mm for all layers (test case 4). The circulation of the working

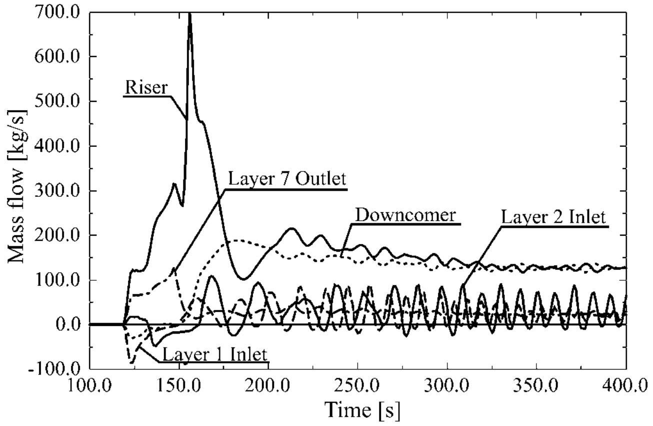


Fig. 13 Mass flow in various tubes of the HRSG with a horizontal tube bank: dynamic behavior with d_a=48.3 mm at all layers (case 4)

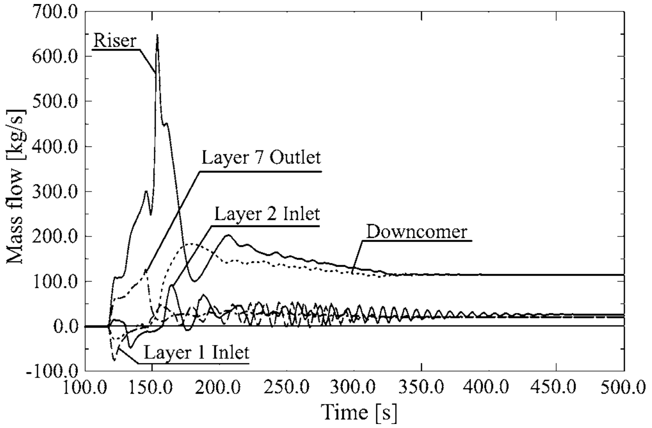


Fig. 14 Mass flow in various tubes of the HRSG with a horizontal tube bank: dynamic behavior with d_a=44.5 mm for the first four layers (case 5)

fluid during the cooling and the first heating phase of the bundled heating surface (up to ~120 s after startup) is ~0. With the beginning of the steam production in the first layer of the heating surface, the volume of the working medium changes rapidly. With the increase of the fluid volume the water-steam mixture will be pushed out at both tube ends. This causes an upward direction of the mass flow in the downcomer as well as in the riser tubes of the boiler. The start of steam production at the other layers of the bundled heating surface is shifted in time. Between the time period of 120–150 s, steam is pushed into the downcomer. With the increasing steam content in the siphon as well in the evaporator tubes, the lifting force reduces the velocity of the water-steam mixture. This results in a stagnation followed by a change of flow direction in the downcomer from upward to downward. Parallel to the change of the flow direction in the downcomer, the flow direction in parts of the evaporator tubes also changes. During this period, the mass flow in the tube network of the evaporator will be let out through the riser to the drum. After 150 s, the mass flow from the drum through the downcomer increases. The mass flow inside the riser after achieving its highest value decreases (the time period between 170 s and 190 s). In the following time period, the mass flow in the downcomer and the riser as well as in the bundled heating surface oscillates. This oscillation does not decay and exists also at steady state.

The oscillation is identified as a density wave oscillation. The phase displacement of the oscillation in the four tube paths of the evaporator is shifted in time. The oscillation amplitude at the inlet of the first four layers is compared to the outlet of the lower heated layers higher. The oscillation is damped during its way through the tube paths from, for example, the inlet of layer 1 to the outlet of the associated layer 7. Therefore, the influence on the mass flow in the downcomer and riser, as well as on the steaming rate, is small.

At the inlet of layers 1–4, the amplitude of the oscillation is high compared to the outlet, and therefore, reverse flow of the working medium occurs inside these tubes for short time periods during the oscillation.

Results of Simulation With an Outer Tube Diameter of d_a =44.5 mm at Layers 1–4. The change of the outer tube diameter at the layers 1–4 results in a homogenization of the heat flow to the single layers of the bundled heating surface, while the total heat flow to the evaporator is approximately constant. The difference of the total heat absorption at steady state between the two test cases is ~0.88% of full load. This leads to essentially the same production rate of steam in both cases.

The heat flow to the first layer of the bundled heating surface at steady state is (in case of an outer tube diameter of 44.5 mm) ~5.8% smaller than in test case 4, whereas at layer 8, the values were ~7.8% higher.

Figure 14 shows the mass flow distribution at selected points of the tube network of the HRSG model with an outer tube diameter of 44.5 mm in the first four layers and a outer tube diameter of 48.3 mm in layers 5–8, which are exposed to a lower heat flux (test case 5). The time evolution of the mass flow during the first 170 s after the start of simulation is similar to that described in test case 4. Differences can be seen during this period only in the absolute values of the mass flows. During the period after the first 170 s, the mass flow starts also to oscillate, but the oscillation amplitude is smaller compared to that in test case 4. During further simulation of the hot startup, the oscillation of the mass flow in the evaporator tubes decreases. Approximately 470 seconds after the start, the density wave oscillation has been eliminated.

The circulating mass flow through the tube network of the HRSG for test case 5 is smaller than the mean value for the mass flow of test case 4. This smaller mass flow results in an earlier beginning of steam production and compared to test case 4 in a slightly higher steam quality at the riser outlet.

Figure 15 shows the development of the total pressure difference for both test cases among the following:

- the drum and the lower header (Δp_{TDL})
- the upper and the lower header (Δp_{TUL})
- the inlet and outlet of the layer 1 (Δp_{TL1}), which is the surface with the highest heat flux

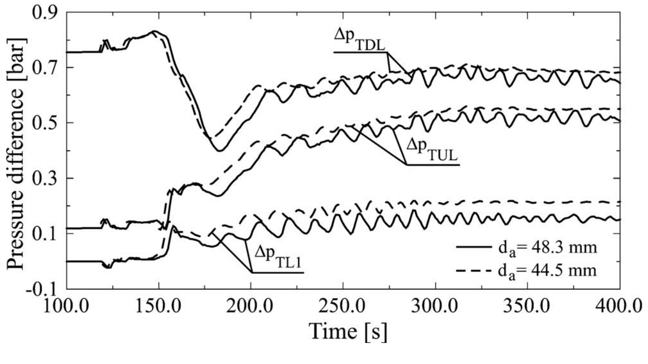


Fig. 15 Comparison between the total pressure differences of the test cases 4 and 5 at all layers

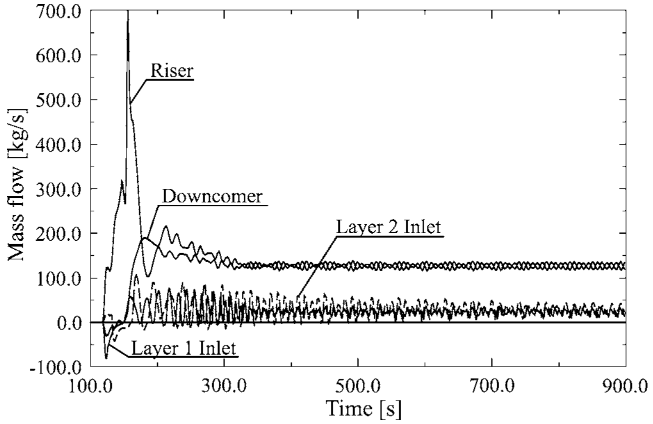


Fig. 16 Mass flow in various tubes of the HRSG with a horizontal tube bank: dynamic behavior with d_a=44.5 mm for the connecting tubes between the header and the first four layers (case 6)

It can be seen that the total pressure differences associated with tubes of 44.5 mm diam are higher than for the tubes of 48.3 mm diam. It seems that a considerable part of the higher total pressure difference between the drum and the lower header is produced in the evaporator tubes with the smaller tube diameter (see Fig. 15, e.g., layer 1).

Results of Simulation With an Outer Tube Diameter of d_a =44.5 mm for Connecting Tubes Between Lower Header and Finned Layer Tubes (With and Without an Orifice). Simulation results for the last two evaporator tube designs are described in the following section. The development of the mass flow for these two test cases takes place during the first period of the hot startup similar to that shown in Fig. 13 for test case 4. Differences are found in the amplitudes and frequencies of the density wave oscillations.

The oscillation for test case 6 appears between 300 s and ~800 s as one with changing amplitude (see Fig. 16). After 800 s, the oscillation shows as one with constant amplitude. The absolute value of the amplitude is smaller than that for the test case 4, and therefore, there is no reverse flow inside the single tubes of the tube bank during the oscillation as compared to test case 4.

The amplitudes and frequencies of the density wave oscillations in test case 7, compared to test case 4, start with smaller values and the amplitude decreases with the progression of the simulation time. Approximately 2350 s after the start of the simulation, the oscillation has ended (Fig. 17). A comparison of test cases 4–7 shows that an increase of the flow resistance at the inlet of the tubes improves the flow stability. The higher flow resistance at the tube inlet increases the single-phase flow resistance, which is in phase with the flow at the tube inlet. The influence of the homogenization of the heat flow to the individual layers of the bundled heating surface, in combination with a additional flow resistance at the tube inlet, obviously improves the stability of the system (see Fig. 14). This effect will be analyzed, in detail, by the authors in further investigations.

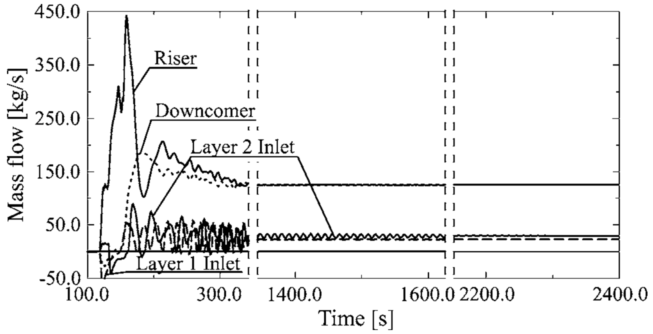


Fig. 17 Mass flow in various tubes of the HRSG with a horizontal tube bank: dynamic behavior with d_a=44.5 mm for the connecting tubes between the header and the first four layers and an additional orifice at the tube inlet (case 7)

Summary

Based on the examples of a natural circulation two-drum heat recovery steam generator and a natural circulation HRSG with a horizontal tube bank it can be shown that design criteria exist which can help to stabilize the flow of the boiler. For the two-drum boiler, the stabilization of the working fluid can be achieved by the arrangement of plain downcomer tubes and, in every case, including cyclones in flow path of the riser tubes stabilizes the circulation of the working medium.

The investigations for the HRSG with a horizontal tube bank show that changes in the bundle geometry can improve the flow stability under operating conditions where density wave oscillations occur. First results of the investigation are presented and discussed. To improve the flow stability, the flow resistance at the tube inlet (single phase flow) of the bundled heating surface should be increased. A faster decay of the density wave oscillations may be also achieved by the homogenization of the heat absorption in the single layers of the bundled heating surface. This provision should be combined with additional flow resistance at the tube inlet.

Nomenclature

A = cross-sectional area (m^2)
d_a = outer diameter (m)
g_x = component of the gravity in direction of the tube axis (m/s^2)
h = spez. enthalpy (J/kg)
h_C = spez. enthalpy of the fluid inside the header (J/kg)
$\dot{m}_0$ = mass flow in the downcomer (kg/s)
$\dot{m}_1$ = mass flow in the lower heated riser system 1 (kg/s)
$\dot{m}_2$ = mass flow in the higher heated riser system 2 (kg/s)
p = pressure (Pa)
P_C = pressure of the fluid inside the header (Pa)
P_D = drum pressure (bar)
Δp_H = pressure difference due to height (static head) (bar)
Δp_{TD} = total pressure difference between bottom drum and drum (bar)
Δp_{TDL} = total pressure difference between drum and lower header (bar)
Δp_{TUL} = total pressure difference between upper and lower header (bar)
Δp_{TL1} = total pressure difference between the inlet and outlet of layer 1 (bar)
$\dot{q}$ = heat flux (W/m^2)
$\dot{q}_1$ = heat flux to the lower heated riser system 1 (kW/m^2)
$\dot{q}_2$ = heat flux to the higher heated riser system 2 (kW/m^2)
t = time (s)
U = perimeter (m)
V = heat absorption ratio
V_C = volume of the fluid inside the header (m^3)
w = velocity (m/s)
x = length (m)

ρ = density $(\mathrm{kg/m^3})$
ρ_C = density of the fluid inside the header $(\mathrm{kg/m^3})$
ζ = pressure loss coefficient

References

[1] Ledinegg, M., 1938, "Instability of Flow During Natural and Forced Circulation," Die Wärme, **61**(48), pp. 891–898.
[2] Bouré, J. A., Bergles, A. E., and Tong, L. S., 1973, "Review of Two-Phase Flow Instability," Nucl. Eng. Des., **25**, pp. 165–192.
[3] Linzer, W., and Walter, H., 2003, "Flow Reversal in Natural Circulation Systems," Appl. Therm. Eng., **23**(18), pp. 2363–2372.
[4] Delhaye, J. M., Giot, M., and Riethmuller, M. L., 1981, *Thermohydraulics of Two-Phase Systems for Industrial Design and Nuclear Engineering*, McGraw-Hill, New York.
[5] Walter, H., 2001, "Modelling and Numerical Simulation of Natural Circulation Steam Generators," Fortschritt-Berichte VDI, Series 6, Report No. 457, VDI-Verlag, Düsseldorf.
[6] Friedel, L., 1979, "Improved Friction Pressure Drop Correlation for Horizontal and Vertical Two-Phase Pipe Flow," European Two-Phase Group Meeting, Ispra, Italy, Paper No. E 2, pp. 1–25.
[7] Patankar, S. V., 1980, *Numerical Heat Transfer and Fluid Flow* (Series in Computational Methods in Mechanics and Thermal Sciences), Hemisphere, Washington, DC.

Evaluation of Interstage Water Injection Effect on Compressor and Engine Performance

I. Roumeliotis[1]
Research Assistant
e-mail: jroume@ltt.ntua.gr

K. Mathioudakis
Associate Professor

Laboratory of Thermal Turbomachines,
National Technical University of Athens,
Iroon Polytechniou 9, Athens 15773, Greece

The present paper examines the effect of water injection at the compressor inlet or between stages, on its operation. A wet compression model coupled with an engine performance model is used. The wet compression model produces the compressor performance map when water is present and consists of a one-dimensional stage stacking model, coupled with a droplet evaporation model. The effect of water injection on overall performance and individual stage operation is examined. The map-generation procedure is embedded in an engine performance model and a study of water injection effect on overall engine performance is undertaken. The possibility to evaluate the effect on various parameters such as power, thermal efficiency, surge margin, as well as the progression of droplets through the stages is demonstrated. The results indicate that water injection causes significant stage rematching, leading the compressor toward stall and that the performance enhancement is greater as the injection point moves towards compressor inlet. [DOI: 10.1115/1.2135823]

Introduction

The increasing use of gas turbines in the power generation industry has created an additional incentive for the further improvement of their performances. In recent years, several techniques have been proposed for gas turbine power and efficiency augmentation, such as steam or water injection into the combustion chamber or the air flow, with much interest in the moisture air cycle [,2] and evaporative gas turbine [3]. The fact that gas turbine output and efficiency drop during high ambient temperature periods, when demand usually increases, has led to the broad application of inlet air cooling [4]. Overspray of water droplets inside the compressor has found an increasing application.

The spray of water inside the compressor in order to achieve power and efficiency augmentation has been proposed from the early years of the gas turbines, by Ægidius Elling [5] and further discussed in order to obtain thrust augmentation in jet engines [6]. A rigorous thermodynamic analysis concerning the thermodynamic of water injection was presented by Hill [7], in which the shift of the compressor characteristics has also been discussed. Further studies of inlet water injection has been undertaken recently by Utamura et al. [8] for a single-shaft engine, Sexton et al. [9,10] for a two-shaft engine, while Meacock et al. [11] presented results for a three-spool engine.

Another approach for power augmentation is interstage injection. Operational aspects were thoroughly discussed by Ingistov [12,13] whereas the effect of interstage injection on compressor and engine performance has been discussed from a thermodynamic point of view by Bagnoli et al. [14], using a commercial program. Also aspects of inlet and interstage injection on the performance of gas turbine has been discussed by Arsen'ev et al. [15].

As presented above, significant attention has been drawn on inlet injection, whereas interstage injection has received little attention, despite the fact that the results presented by Arsen'ev et al. [15] indicated that interstage injection may cause greater performance enhancement than inlet injection.

Another interesting aspect of water injection, which needs further investigation, is the effect of stage rematching to the overall compressor operating limit and on the engine operation envelope, as surge margin is a significant factor of engine development and operation.

In the present paper, a method of studying the effects of water injection on compressor and engine performance is presented. Water-droplet evaporation through the stages of a multistage compressor is modeled in conjunction to a stage-stacking model for map generation. An innovative approach for coupling this procedure with a zero-order engine model is proposed, and effects on overall performance are studied.

Compressor Model

In order to evaluate compressor performance with water injection at the inlet or intermediate positions a "stage-stacking" approach coupled with a droplet model has been chosen. The effect of the presence of water on the operation of each individual stage is estimated, and the overall performance is deducted by "stacking" the individual performances.

The compressor map for various operating speeds with or without water injection can thus be derived. Not only are overall performance quantities produced, but also operational characteristics, such as the surge line. In the following, the model for droplet evaporation is first described, followed by a description of its incorporation to the stage-stacking procedure.

Droplet Model. The droplet model adopted for the present study is the model described by Spalding [16]. The droplets are assumed to be entrained in the flow, so there is no velocity slip between them and the gaseous medium. According to White and Meacock [17], the droplet diameter must be up to 5 μm for the no-slip assumption to be valid.

The droplet temperature is determined by the droplet energy balance, where the droplet energy change equals the conductive heat transfer and the latent heat transfer

$$\frac{dT_d}{dt} = -\frac{3(\dot{Q}_0 + jh_{fg})}{\rho_{\text{liq}} r_d c_{p\text{liq}}} \tag{1}$$

The vapor mass flux may be calculated using to the following relation:

---

[1] Corresponding author.

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the Journal of Engineering for Gas Turbines and Power. Manuscript received October 1, 2003; final manuscript received March 1, 2004. IGTI Review Chair: K. C. Hall. Paper presented at the ASME Turbo Expo 2005: Land, Sea, and Air, Reno, NV, June 6–9, 2005, Paper No. GT2005-68698.

$$j = \frac{\rho_g D_{\mathrm{H_2Oair}}}{r_d} \ln\left(\frac{m_{vg} - 1}{m_{vs} - 1}\right) \tag{2}$$

The pressure on the droplet surface is assumed to be equal to the saturation pressure.

The diffusion coefficient can be calculated through Eq. (3), by the formula proposed by Marrero and Mason [18], which suits the temperature range for the case of compressor

$$D_{\mathrm{H_2Oair}}(m^2/s) = \begin{cases} 1.87 \times 10^{-10} \dfrac{T_g^{2.072}}{p_g}, & 280\ \mathrm{K} < T_g < 450\ \mathrm{K} \\ 2.75 \times 10^{-9} \dfrac{T_g^{1.632}}{p_g}, & 450\ \mathrm{K} < T_g < 1070\ \mathrm{K} \end{cases} \tag{3}$$

p_g is the air pressure in at and T_g the air temperature in Kelvin.

The heat flux through the gas phase may be calculated according to energy and mass conservation using the relation

$$\dot{Q}_0 = \frac{j c_{pv}(T_g - T_s)}{1 - \exp\left(\dfrac{r_d j c_{pv}}{k}\right)} \tag{4}$$

The droplet radius change is calculated using the mass conservation

$$\frac{dr_d}{dt} = -\frac{j}{\rho_{\mathrm{liq}}} = \frac{\rho_g D_{\mathrm{H_2O\ air}}}{r_d \rho_{\mathrm{liq}}} \ln\left(\frac{m_{vs} - 1}{m_{vg} - 1}\right) \tag{5}$$

Stage-Stacking Model. The overall performance of a multistage axial compressor depends on the performance of its constituent stages. The performance of a single stage can be presented in terms of nondimensional coefficients, namely, flow coefficient Φ, pressure coefficient Ψ, and efficiency η. From individual stage characteristics, the overall compressor performance for desired operating points may by evaluated using the stage-stacking technique. The technique has been developed a while ago, and descriptions may be found in a number of sources (e.g. [19]).

The particular stage-stacking model used in the presented study is the one introduced by Mathioudakis and Stamatis [20]. It incorporates an adaptivity feature allowing the derivation of individual stage characteristics, which optimally reproduce a given overall map.

The model uses generalized characteristics, which interrelate the normalized parameters Φ/Φ_{ref}, Ψ/Ψ_{ref}, and η/η_{ref}. Using the generalized characteristic curve and the values of Φ_{ref}, Ψ_{ref}, and η_{ref} for each stage, the characteristics of the different stages are deduced. Using the non-dimensional parameters in conjunction with the geometric data, the entropy and enthalpy increase through the stage for the gaseous mixture are calculated, along with the properties, namely, pressure temperature (stagnation, static) and velocity at each stage.

In order to incorporate the process of water evaporation into a stage stacking method, there are some important factors to be addressed.

1. The definition of stagnation magnitudes, due to the occurrence of heat transfer, while the mixture is not in thermal equilibrium must be reexamined. In order to avoid physical inconsistencies, calculations are performed employing the static properties, where stagnation properties change is accounted for through the total enthalpy rise in individual stages. The stagnation properties are finally calculated at the compressor exit, assuming no water presence, which is true for most actual application test cases.
2. Droplet evaporation is a time-variant process, which means that time should be incorporated into the stage-stacking method.
3. Evaporation is an irreversible process leading to entropy increase so the losses due to the irreversible nature of evaporation should be incorporated into the model, as White and Meacock [17] have suggested, according to the relation presented by Young [21].
4. According to Csanady [22], the mixture composition has no significant effect on stage characteristics; thus, the stage characteristics are assumed unchanged due to vapor presence. In the case of water injection, no experimental data for stage characteristic shift have been presented. With the assumption that no losses of a hydrodynamic or mechanical nature occur (film creation, rotor braking), the compressor characteristics are assumed unchanged because of water injection.

The time dimension entering the droplet evaporation model is linked to the stage-stacking technique by using rotor and stator lengths and calculating the mean residence time of the droplets inside each blade row using flow velocity. As a first approximation, the axial velocity is considered constant along a stage.

The shift of the stage operation point due to water droplet presence is evaluated by calculating the volume occupied by the droplets and correcting axial velocity by the corresponding blockade factor. For this purpose the actual mass and volume of droplets should be known at the inlet and outlet of each stage.

The mass of water is estimated via the following:

$$q_{\mathrm{liq}} = \rho_{\mathrm{liq}} \cdot \frac{4 \cdot \pi}{3} r_d^3 \cdot n_{dr} \tag{6}$$

In the present model the number of droplets n_{dr} is assumed to remain unchanged through the stages, and it is the droplet diameter that is reduced until full evaporation.

The flow parameter determines a unique stage efficiency η and pressure-rise coefficient Ψ. Using these parameters, the total enthalpy and entropy increase of the gaseous mixture is calculated, assuming that no evaporation occurs. The static enthalpy and entropy increase is also calculated, with the assumption of constant axial velocity. It is assumed that all of the energy required to evaporate the water is absorbed from the gaseous phase. The entropy increase, representing the aerodynamical losses, is distributed equally between the rotor and the stator.

The conditions of the flow at the rotor outlet are first calculated as if no evaporation occurs, according to the calculated entropy and enthalpy increase of the mixture. After calculation of the mixture properties, an iterative procedure is established in order to calculate the actual conditions when evaporation occurs. The droplet model is coupled, with a set of equations representing the static enthalpy increase due to compression and the static entropy increase due to evaporation, according to the residence time in the rotor.

The same procedure is followed for the stator, with the addition that a new velocity is calculated according to the exit angle and annulus. The exit angle of the stage is assumed to be the inlet angle of the subsequent stage.

Test Case for Wet Compression. In order to demonstrate the effects predicted by the present method onto a compressor map, a multistage axial compressor has been used as a test case. The 15-stage compressor reported by Tsalavoutas et al. [23] has been used. The dry-stage characteristics have already been derived in [23] so that the overall map of that compressor could be reproduced. In Fig. 1 an example of a set of dry and wet characteristics with water injection at the inlet is presented, for 15 °C and saturated air at the compressor inlet. The total inlet flow (gaseous and liquid) is used in the abscissa.

The observed behavior is similar to previously reported results by Hill [7] and White and Meacock [17] for the case of inlet water injection. The gaseous inlet mass flow is increasing as the injected

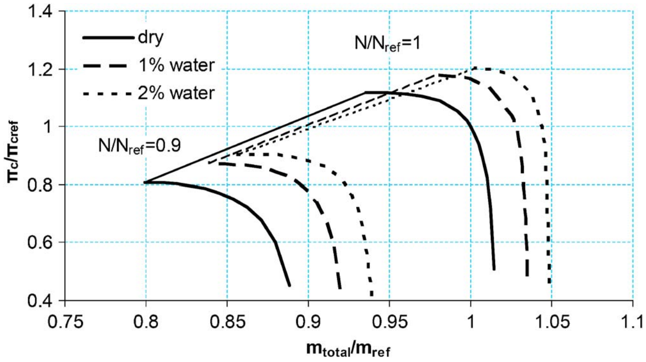


Fig. 1 Compressor map for inlet injection

percentage of water increases. It must be noted here that in one case [24], the opposite effect has been reported. The authors attributed their observation to the particular design of the compressor.

However, the inlet gaseous mass flow increase rate is lower as the quantity of injected water is increasing. The increase of pressure ratio results in a degradation of the aerodynamic performance, as the compressor is not designed for water injection; thus, the stages are not operating anymore at the optimum operating point. In order to demonstrate the change of aerodynamic performance as the evaporation is completed inside the compressor, the "equivalent" polytropic efficiency of the air, as proposed by White and Meacock [17] is adopted

$$\eta_a = \frac{1}{1 + \Delta s_g/(R_g \ln p_{t_2}/p_{t_1})} \tag{7}$$

This can be used in order to demonstrate the aerodynamic effect of water injection for the case of interstage injection. The equivalent polytropic efficiency is presented in Fig. 2.

Interstage Injection. The next step is the evaluation of interstage water injection at the compressor map. In this case, the inlet air mass flow is used in the abscissa. Thus the presented mass flow is increased by the amount of injected water at the rear of the compressor. As seen in Fig. 3, the same behavior as for inlet injection is observed for the case of interstage water injection. The map presented is for water injection at the inlet of the third stage. The inlet gaseous mass flow is increased, and the aerodynamic efficiency is decreased.

As one can see in Fig. 4, as water is injected further downstream of the compressor inlet the net increase in mass flow is reduced. On the other hand, there is a slight improvement in the equivalent aerodynamic performance as water is injected farther from the inlet. The change is smaller when the injection takes place at the second or third stage for the high-speed characteristic, due to low evaporation rates at the first stages.

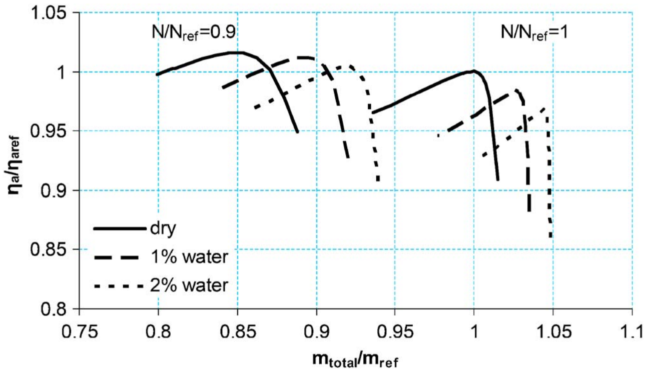


Fig. 2 Equivalent polytropic efficiency for inlet injection

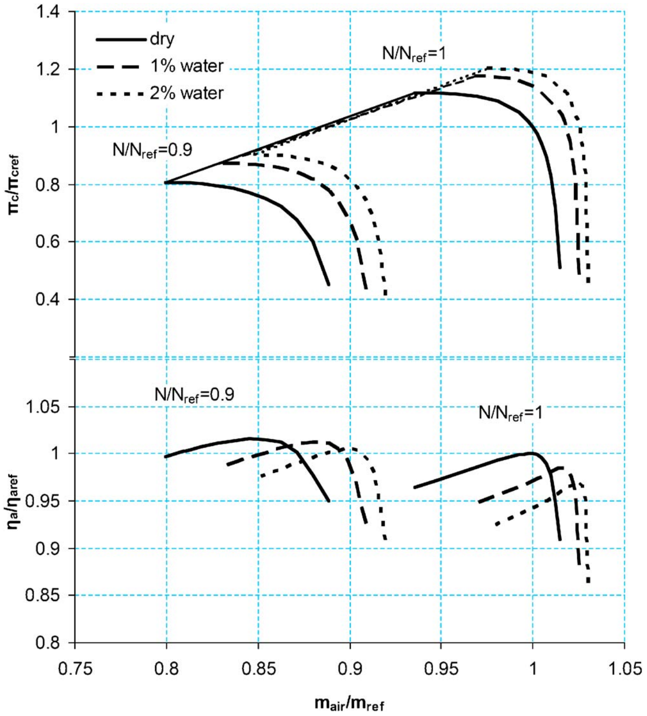


Fig. 3 Compressor map for third-stage injection

Operating Limits. Water injection alters the overall compressor characteristics, as stage rematching occurs. These alternations may also alter the surge limit. Concerning the map of the selected compressor, the criterion that gave results in good agreement with available data at high rotational speeds is that stall occurs at the

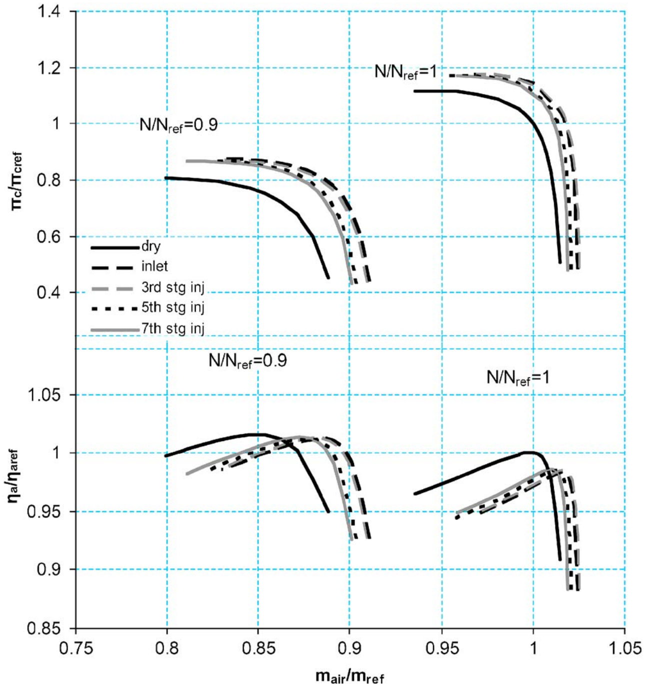


Fig. 4 Compressor map for various injection positions, for 1% of injected water

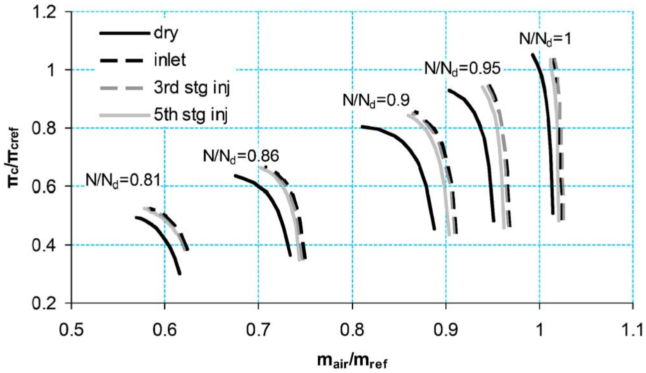


Fig. 5 Compressor map, with established stall criterion for 1% injected water

maximum pressure-rise coefficient for any individual stage. This means that the stage is considered in stall when $d\Psi/d\Phi=0$, for the specific characteristic, and at that point the limits of the compressor are established. Using this criterion, the operation of the compressor with water injection becomes more limited. As with water injection the rear stages are working toward greater Ψ, thus not gaining the significant margin presented in the previous figures. This criterion will be used during the following study, and the corresponding full map of the compressor with 1% of water injected at compressor inlet, and at the inlet of stages 3 and 5, is presented in Fig. 5. For $N/N_d<0.9$, the IGVs and VSVs are closing according to the data presented by Carchedi et al. [25].

Engine Model

In order to investigate the effect of inlet and interstage water injection on the performance of a gas turbine engine, the method described above is coupled to an engine performance model. The model described in [26] has been used. It can incorporate the presence of air humidity as well as the injection of water or steam at any station along the gas path, assuming that evaporation takes place in mixing chambers. That version, though, handled injected water as if thermodynamic equilibrium is reached at once, without taking into account evaporation dynamics.

Water injection alters the characteristics of the compressor, depending on the position and amount of injected water. In order to have the possibility to incorporate injection effects as accurately as possible and to understand individual stage behavior, the stage-stacking model was incorporated into the engine model instead of using the map. A kind of one-dimensional (1D) zooming is thus applied.

The way of coupling the two codes is shown in Fig. 6. The error variables e_{outi} that are zeroed by the model in order to balance the equations are for the particular example of a single-shaft engine that is used here as a test case. It is recalled that the model has the ability to model up to three spool engines with or without power turbine and for a variety of control variables as presented in [26].

The procedure is as follows: first the stage-stacking model is used to create a compressor map for the particular inlet conditions, water injection amount, and position. This map is used to derive an estimated mass flow rate m_c for a guessed pressure ratio π_c. Using this mass flow, the stage stacking is applied to obtain the properties of the flow at the compressor exit (e.g., compressor work, pressure, temperature, airflow, water flow), as shown in Fig. 7. The whole procedure is iterated until the model converges.

The test case of a single-shaft gas turbine was used for the application of the method. The single-shaft industrial gas turbine modeled in [27] has been used.

Droplet Evaporation. As seen from the equations describing the Spalding model, evaporation rate is a function of droplet diameter, droplet conditions, and local gas conditions. Evaporation quantity is a function of the residence time in the compressor. For the specific compressor studied, residence time was about 6 ms. In the present study, the gas phase was assumed homogeneous at any axial position except in the droplet boundary layer, where the local water-to-air ratio depends on the droplet temperature.

During wet compression, the gas properties change throughout the compressor, along with the operation of the compressor, thus influencing the droplet evaporation rate. For example in the case of 5% of water injected at the first compressor stage, the exit temperature can be decreased by 30% or even more, depending on the engine operating point. Thus, the full evaporation position is moving toward the rear of the compressor as the water injection fraction increases.

The results for the position of evaporation completion through the compressor, presented in Fig. 8, are in agreement with the results for droplets of 5 μm diam presented by Loebig et al. [28], who use a more complex three-dimensional (3D) analysis, for a compressor with similar length and temperature rise characteristic.

Stage Rematching Due to Droplet Evaporation. In order to examine the stage rematching caused by droplet evaporation, the injection was assumed to take place at the inlet of the first stage; thus, no evaporation occurs before the first stage inlet. This is because inlet injection alters the corrected speed due to inlet cooling; thus, the droplet evaporation effect inside the compressor cannot be isolated. In Fig. 9, the generalized $\Phi-\Psi$ characteristic is presented along with the operating points of the stages 1 and 15, for various percentages of water injection. The case of 2% of water injection is limiting for the specific compressor, at the specific operation point, since the 15th stage operation approaches the stall criterion selected.

Loading reduces for the front stages due to the increased air mass flow rate, with the additional liquid mass flow. Further downstream, the cooling effect and the higher achieved pressure ratios result in higher density, thus lower Φ. This drives the rear stages to operate closer to stall. Increasing the amount of injected water, the operation of the stages moves closer to stall. This is indicated by the shift of the Φ/Φ_{dry}, presented in Fig. 10, where it is shown that rear stages work at lower flow coefficient as the amount of injected water increases.

The injection position has an impact on the stage loading throughout the engine, as shown in Fig. 11, for injection at the inlet of the first, third, and fifth stages. The trend for stage rematching presented for the case of interstage injection accounting for the evaporation process inside the compressor is similar to the one presented in [14], where thermodynamic equilibrium was assumed.

An interesting aspect in engine operation is the surge margin. The cooling effect has a significant effect on the surge margin, since the pressure ratio increases, and as presented above the rear stages work with higher pressure coefficient.

In Fig. 12 the compressor operating point for constant TIT is presented, along with the corresponding calculated speedline, for water injection at the first stage. The surge margin reduces even with low injection quantities. The same observations concerning the trend of the surge margin have been presented by Ludorf et al. [29], for the case of high rotating speeds. The study by Horlock [30] indicates that the later stages are expected to work at a higher pressure coefficient with water injection.

For water injection in the third stage, no significant changes in the surge margin were evident. Moving the water injection position toward the end, the surge margin, according to the criterion established above appears to reduce. The effect is increasing for higher water injection flows. In Fig. 13 the surge margin for the specific engine operating point is presented. The surge margin definition used is

$$\text{Surge margin}(\%) = \frac{\pi_{c\text{surge}} - \pi_{c\text{working}}}{\pi_{c\text{working}}} \cdot 100 \tag{8}$$

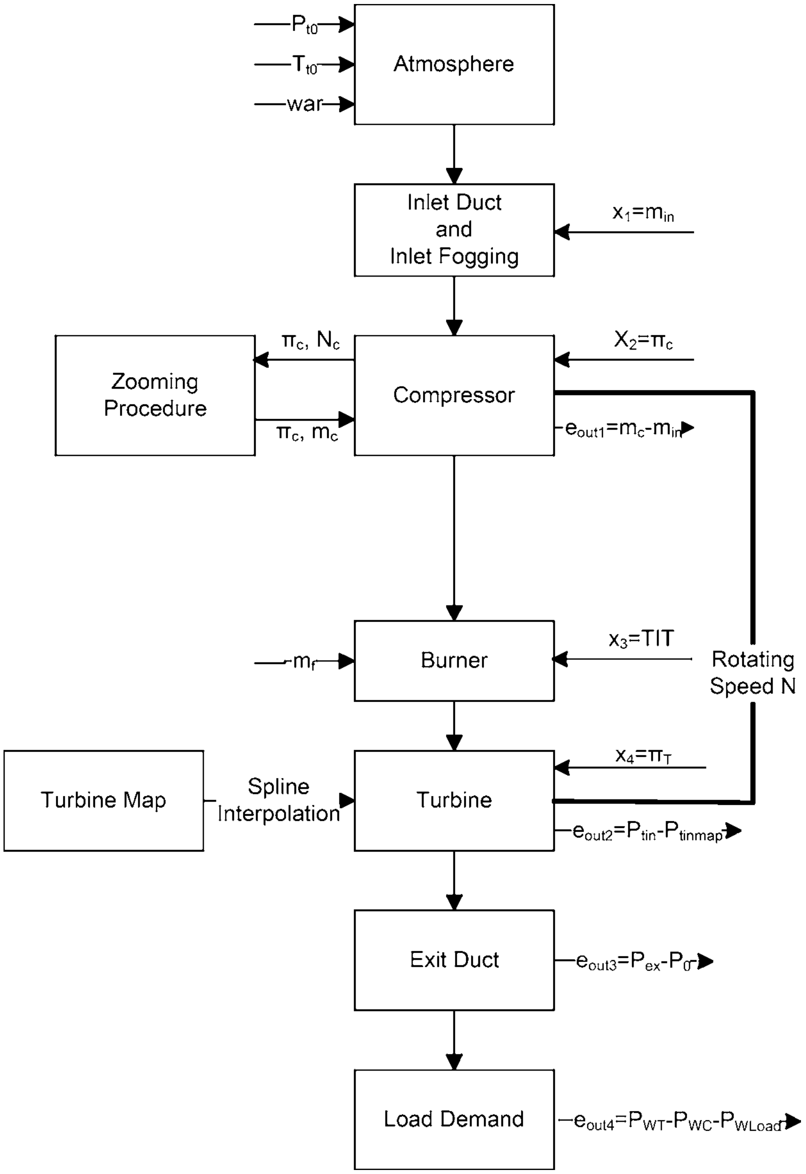


Fig. 6 Coupling of stage-stacking model and performance model

Effect on Overall Performance. The impact of the water injection position on the performance characteristics of the engine is now examined. The behavior of the engine performance concerning the effect of injection position was examined for ambient temperature of 30 °C with constant TIT ($TIT/TIT_{ref}=0.95$). The injection positions are the inlet of the first, third, and fifth stage.

The expected positive effect of water injection in the thermal efficiency is observed as presented in Fig. 14. The thermal efficiency gain is lower as the injection point moves toward the rear of the compressor, although it is rather high for every injection position. The same applies to the power of the engine, as the load gain is reduced applying water injection at later stages, although the load reduction is rather small for injection at the third stage.

Another interesting operational aspect is the effect on exhaust gas temperature and exhaust mass flow, quantities that are important for operation in combined cycles. The exhaust gas temperature has a small decrease while the mass increases, as shown in Fig. 15. In both cases, the more significant effect is caused by injection at the first-stage inlet.

The size of droplet diameter is also a significant factor since smaller droplets keep the conditions during compression closer to thermodynamic equilibrium; thus, the losses due to irreversible process of evaporation are lower. In Fig. 16, the thermal efficiency variation for first-stage injection, for 5 and 2μm diam droplets is presented.

Arsen'ev et al. [15] have presented results indicating that the interstage injection may be more beneficial than inlet injection, when the injection is realized at an optimal position inside the compressor. These conclusions were drawn, taking into consideration mechanical and hydrodynamic losses, which were greatly dependent on the injection position, as in the case of third-stage injection the hydrodynamic losses were half of those for inlet injection, and in the case of fifth-stage injection, the losses were 20% of the losses for inlet injection.

This indicates that experimental results concerning mechanical losses are needed, as their effect may be significant. As the spraying technology is advanced, the no-slip assumption becomes more valid and the results presented here indicate that, in the case that no additional losses due to mechanical and hydrodynamic reason are presented, the injection is more positive as the point of injection is moving toward the inlet of the engine, for ambient temperatures where overspray is usually used.

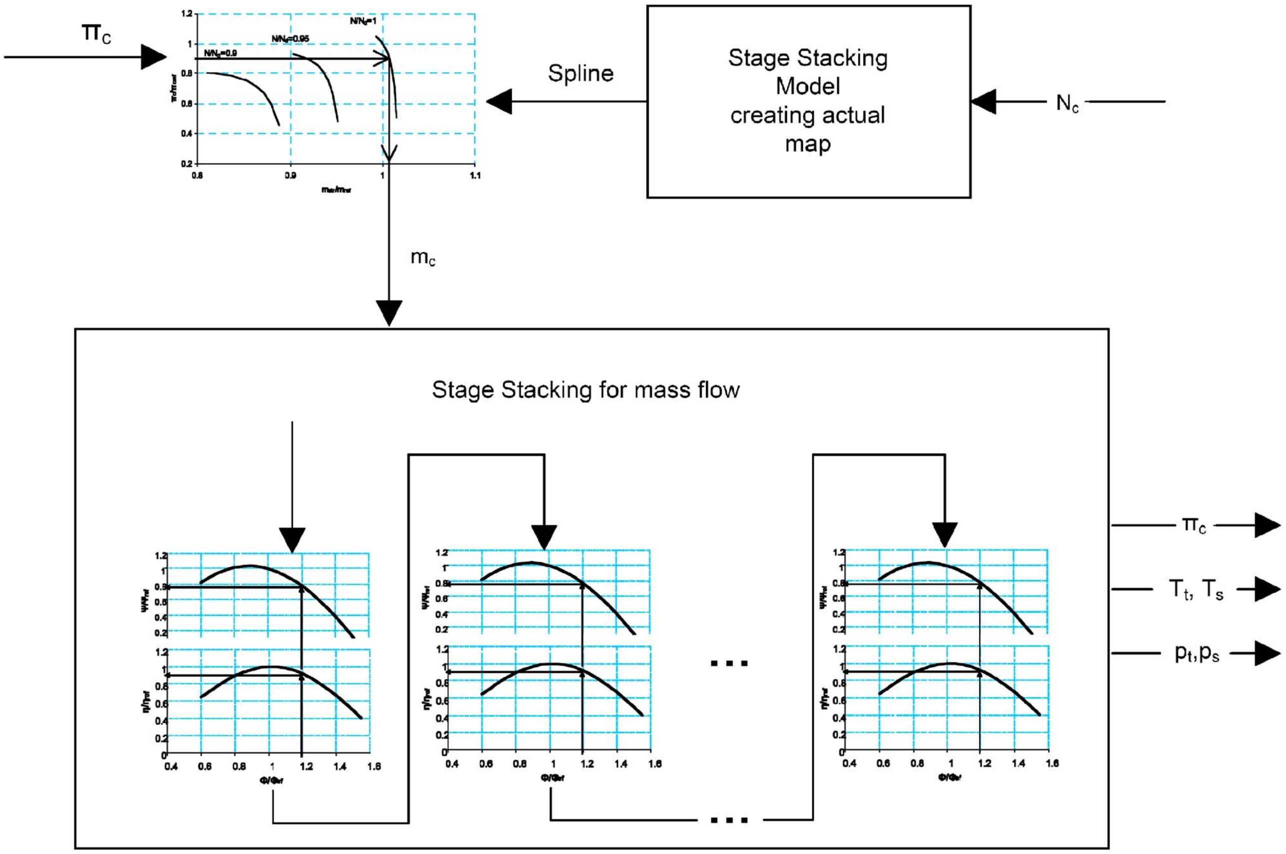


Fig. 7 Zooming procedure

On the other hand, interstage injection has the advantage that it can be used independently of the ambient temperature, for suitable injection points. In this case the injection point and the lower air temperature that interstage injection should be used depend mainly on the compressor characteristics and specifically on the temperature increase that can be achieved at the front stages, along with the anti-icing procedure of the specific engine. From this point of view, interstage injection may be of practical interest.

Conclusions

A method to evaluate the effect of water-injection at the inlet or between stages of a multistage axial compressor has been presented. The method was incorporated into an engine model, allowing the evaluation of water injection effects on overall performance, with high flexibility concerning the engine configuration and control variable.

The behavior of the compressor was studied, and the results indicated that water injection at various positions shifts the characteristics to higher mass flow ratio and compression ratio. However the individual stages are working off design, where the first stages are unloading and the rear stages are shifted closer to stall.

The stage rematching, leading the compressor closer to stall, is bounding the benefits of water injection, as there is a severe reduction on surge margin. The decrease of surge margin is more significant as the injection position is moving toward the rear stages of the compressor, although the difference is small.

A marginal thermal efficiency increase has been predicted, and the effect of interstage injection seems to lower the water-

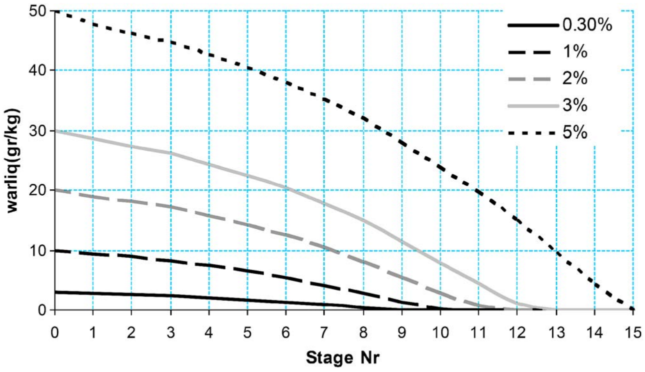


Fig. 8 Evaporative profile inside compressor for 5 μm diam droplets

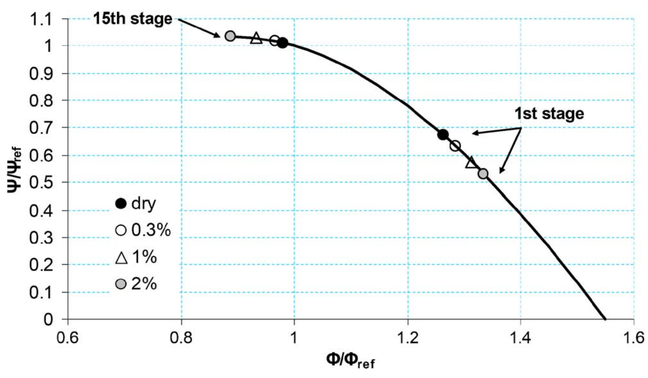


Fig. 9 Stages 1 and 15 operating point shift for various injected quantities

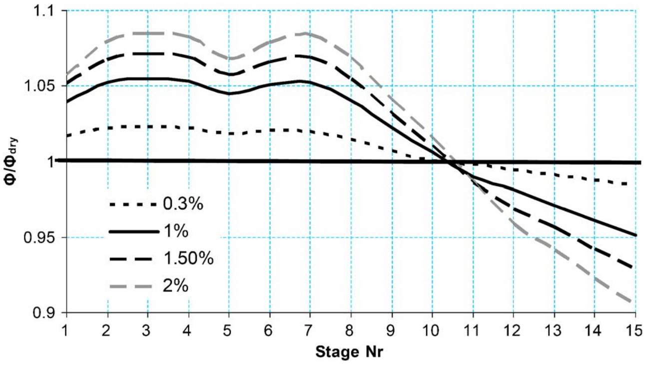


Fig. 10 Normalized flow coefficient for First-stage injection

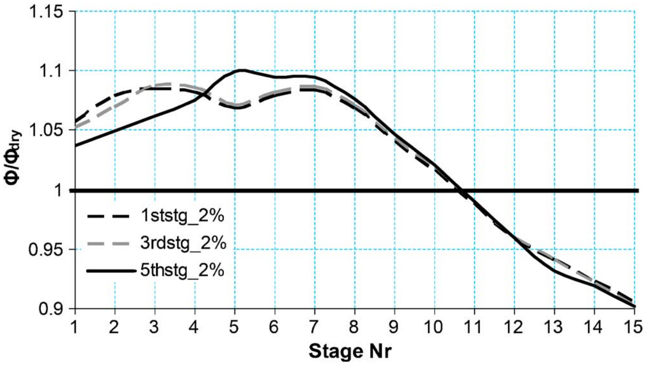


Fig. 11 Normalized flow coefficient for interstage injection

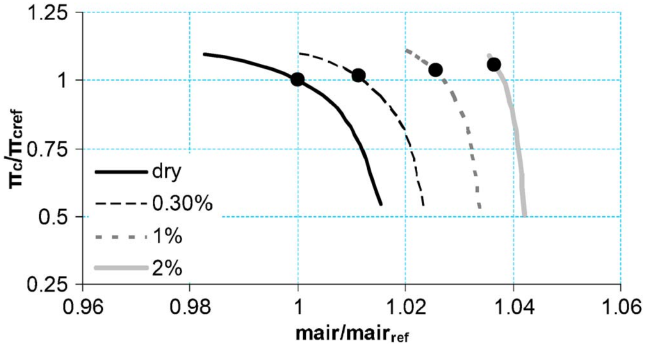


Fig. 12 Compressor operating points for First-stage injection

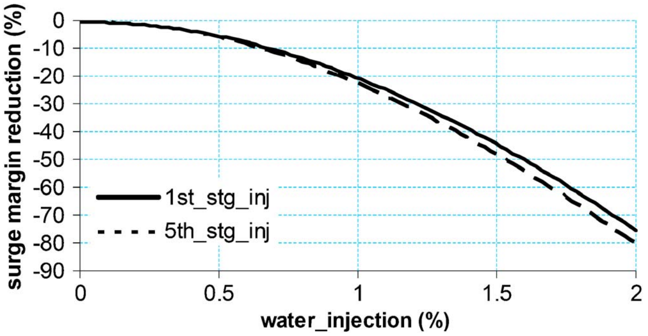


Fig. 13 Surge margin for First- and Fifth-stage injection

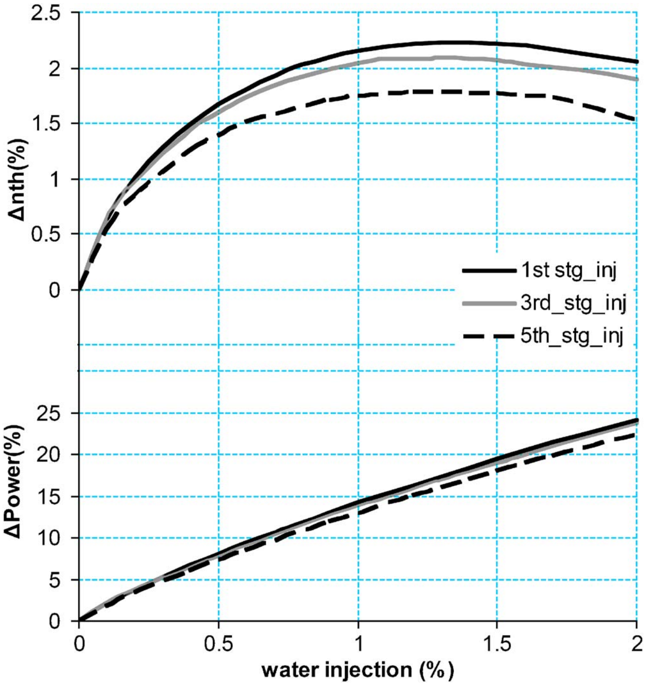


Fig. 14 Gas turbine thermal efficiency and load variation

injection benefits on efficiency, although for the case of third-stage injection, the differences are relatively small. A substantial power boost has been predicted with water injection, the gain being more important as the injection position moves toward the compressor inlet.

The results concerning the significant stage rematching due to water injection indicate that the behavior for specific type of compressor may be different with wet compression; thus, in order to

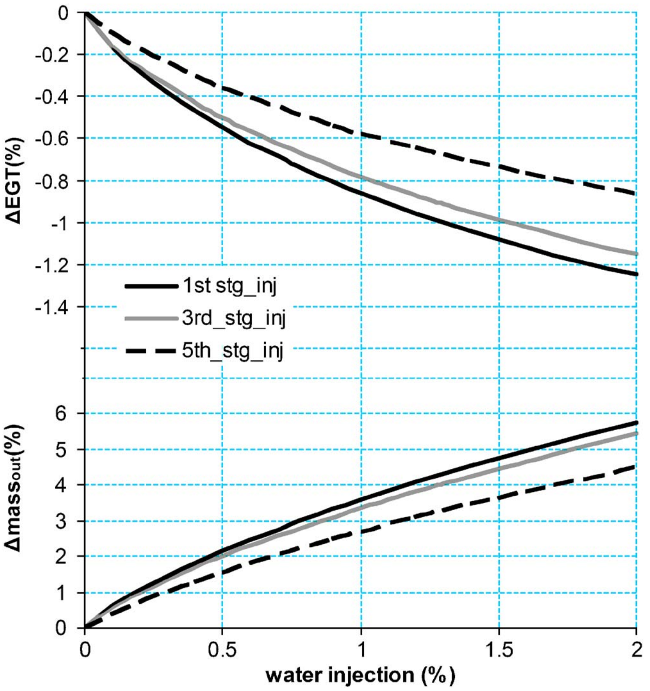


Fig. 15 EGT and exhaust mass variation

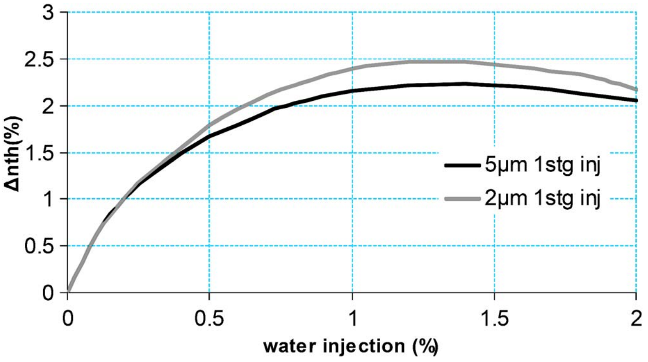


Fig. 16 Thermal efficiency for 2 μm and 5 μm diam droplets

obtain the full benefits of water injection the calculations should be realized with respect to the specific or similar compressor stage characteristics and geometry.

Acknowledgment

The work for this paper has been carried out in the frame of HERAKLITOS program. The program is co-funded by the European Social Fund (75%) and National Resources (25%). The assistance of G. Sieros in part of the work is also acknowledged.

Nomenclature

c_p = isobaric specific heat capacity
$D_{\mathrm{H_2O_{air}}}$ = diffusion coefficient
e_{outi} = error function
h = specific enthalpy
h_{fg} = latent heat of vaporization
IGVs = inlet guide vanes
j = vapor mass flux
k = thermal conductivity
m = mass concentration
n_{dr} = droplet number
p = pressure
Q_0 = conductive heat flux
r_d = droplet radius
T = temperature
TiT = turbine inlet temperature
V_a = axial velocity
VSVs = variable stator vanes
$\eta = \Delta h_{tis}/\Delta h_t$ = stage efficiency
π_c = compressor pressure ratio
ρ = density
$\Phi = V_a/U$ = mass flow coefficient
$\Psi = \Delta h_{tis}/U^2$ = pressure rise coefficient
x_i = guessed variable

Subscripts

g = gaseous mixture
liq = liquid water
s = droplet surface
t = stagnation magnitude
v = vapor water

References

[1] Utamura, M., Takehara, I., and Karasawa, H., 1998, "MAT, A Novel, Open Cycle Gas Turbine for Power Augmentation," Energy Convers. Manage., **39**(16–18), pp. 1631–1642.
[2] Utamura, M., Takehara, I., Horii, N., and Kuwahara, T., 1997, "A New Gas Turbine Cycle for Economical Power Boosting," ASME Paper No. 97-AA-142.
[3] Horlock, J. H., 1997, "The Evaporative Gas Turbine [EGT] Cycle," ASME Paper No. 97-GT-408.
[4] Meher Homji, C. B., and Mee, T. R., III, 2000, "Inlet Fogging of Gas Turbine Engines, Part A: Theory, Psychrometrics, and Fog Generation," ASME Paper No. 2000-GT-307.
[5] Bakken, L. E., Jordal, K., Syverud, E., and Veer, T., 2004 "Centenary of the First Gas Turbine to Given Net Power Output: A Tribure to Ægidius Elling," ASME Paper No. GT2004-53211.
[6] Trout, A. M., 1950, "Theoretical Turbojet Thrust Augmentation by Evaporation of Water During Compression as Determined by Use of a Mollier Diagram," NACA TN 2104, June.
[7] Hill, P. G., 1963, "Aerodynamic and Thermodynamic Effects of Coolant Injection on Axial Compressors," Aeronaut. Q., **14**, pp. 333–348.
[8] Utamura, M., Kuwahara, T., Murata, H., and Horii, N., 1999, "Effect of Intensive Evaporative Cooling on Performance Characteristics of Land Based Gas Turbine," *Proc. International Joint Power Generation Conference*, ASME, NY, Vol. **2** - Power, PWR Vol. **34**, pp. 321–328.
[9] Sexton, M. R., Urbach, H. B., and Knauss, D. T., 1998, "Evaporative Cooling for NOx Suppression and Enhanced Engine Performance for Naval Gas Turbine Propulsion Plants," ASME Paper No. 98-GT-332.
[10] Sexton, W. R., and Sexton, M. R., 2003, "The Effect of Wet Compression on Gas Turbine Operating Performance," ASME Paper No. GT2003-38045.
[11] Meacock, A. J., and White, A. J., 2004, "The Effect of Water Injection on Multi-Spool Gas Turbine Behaviour," ASME Paper No. GT2004-53320.
[12] Ingistov, S., 2001, "Interstage Injection System for Heavy Duty Industrial Gas Turbine Model 7EA," ASME Paper No. 2001-GT-0407.
[13] Ingistov, S., 2002, "Interstage Injection Into Axial Compressor Gas Turbine Model 7EA, Part 2," ASME Paper No. GT2002-30656.
[14] Bagnoli, M. Bianchi, M., Melino, F., Peretto, A., Spina, P. R., Bhargava, R., and Ingistov, S., 2004, "A Parametric Study of Interstage Injection on GE Frame 7EA Gas Turbine, ASME Paper No. GT2004-53042.
[15] Arsen'ev, L. V., and Berkovich, A. L., 1996, "The Parameters of Gas Turbine Units With Water Injected Into the Compressor," Therm. Eng., **43**(6), pp. 461–465.
[16] Spaling, D. B., 1979, *Combustor and Mass Transfer*, Pergamon Press, New York, pp. 82–99.
[17] White, A. J., and Meacock, A. J., 2003, "An Evaluation of the Effects of Water Injection on Compressor Performance," ASME Paper No. GT-2003-38237.
[18] Marrero, T. R., and Mason, E. A., 1972, "Gaseous Diffusion Coefficients," J. Phys. Chem. Ref. Data **1**, pp. 3–118.
[19] Robbins, W. H., and Dugan, J. F., 1965, "Prediction of Off Design Performance of Multistage Compressors," "*Aerodynamic Design of Axial Flow Compressors*," NASA SP 36, I. H. Jonsen and R. O. Bullock, eds., NASA, WA, DC, pp. 297–310.
[20] Mathioudakis, K., and Stamatis, A., 1994, "Compressor Fault Identification From Overall Performance Data Based on Adaptive Stage Stacking," ASME J. Eng. Gas Turbines Power, **116**, pp. 156–164.
[21] Young, J. B., 1995, "The Fundamental Equations of Gas Droplet Multiphase Flow," Int. J. Multiphase Flow, **21**(2), pp. 175–191.
[22] Csanady, G. T., 1964, *Theory of Turbomachines*, McGraw-Hill, NY.
[23] Tsalavoutas A., Stamatis A., and Mathioudakis K., 1994, "Derivation of Compressor Stage Characteristics, for Accurate Overall Perfrormance Map Prediction" ASME Paper No. 94-GT-372.
[24] Cataldi, G., Güntner, H., Matz, C., McKay, T., Hoffmann, J., Nemet, A., Lecheler, S., and Braun, J., 2004, "Influence of High Fogging Systems on Gas Turbine Engine Operation and Performance," ASME Paper No. GT2004-53788.
[25] Carchedi F., and Wood G. R., 1982, "Design and Development of a 12:1 Pressure Ratio Compressor for the Ruston 6-MW Gas Turbine," ASME Paper No. 82-GT-20.
[26] Roumeliotis I., Mathioudakis K., and Aretakis N., 2003, "Performance Analysis of Twin Spool Water Injected Gas Turbines Using Adaptive Modeling," ASME Paper No. GT2003-38516.
[27] Stamatis, A., Mathioudakis, K., Papailiou, D. K., 1990 "Adaptive Simulation of Gas Turbine Performance," ASME J. Eng. Gas Turbines Power, **112**, pp. 168–175.
[28] Loebig J, Vittal B., and Booher M., 1998, "Numerical Simulation of Water/Methanol Evaporation in an Axial Flow Gas Turbine Compressor," AIAA Paper No. AIAA-98-3559.
[29] Ludorf, R. K., Elder R. L., Tronbøl T. H., and Overli J., 1995, "Stage Re-Matching as a Result of Droplet Evaporation in a Compressor," ASME Paper No. 95-GT-194.
[30] Horlock J. H., 2001, "Compressor Performance With Water Injection," ASME Paper No. 2001-GT-343.

Tommy J. George

M.-H. Herman Shen[1]
e-mail: shen.1@osu.edu

Department of Aerospace Engineering and
Aviation,
The Ohio State University,
Columbus, OH 43210

Theodore Nicholas

Charles J. Cross

Air Force Research Laboratory,
Wright-Patterson AFB, OH 45433

A New Multiaxial Fatigue Testing Method for Variable-Amplitude Loading and Stress Ratio

A new vibration-based multiaxial fatigue testing methodology for assessing high-cycle turbine engine material fatigue strength at various stress ratios is presented. The idea is to accumulate fatigue energy on a base-excited plate specimen at high-frequency resonant modes and to complete a fatigue test in a much more efficient way at very low cost. The methodology consists of (1) a topological design procedure, incorporating a finite element model, to characterize the shape of the specimens for ensuring the required stress state/pattern, (2) a vibration feedback empirical procedure for achieving the high-cycle fatigue experiments with variable-amplitude loading, and finally (3) a yielding procedure for achieving various uniaxial stress ratios. The performance of the methodology is demonstrated by the experimental results from mild steel, 6061-T6 aluminum, and Ti-6Al-4V plate specimens subjected to fully reversed bending for both uniaxial and biaxial stress states. [DOI: 10.1115/1.1788687]

1 Introduction

Due to the severe operational environment, gas turbine engine structural components are usually required to be designed using lifetime failure-free criterion, with the use of tools such as the Goodman diagram [1] and the modified Goodman diagram [2], which are usually constructed using uniaxial fatigue data. The Goodman diagram is a plot of alternating stress versus mean stress and represents the fatigue properties of a given material for a given number of cycles. A typical Goodman diagram for the titanium alloy Ti-6Al-4V is shown in Fig. 1 [3].

Typical uniaxial fatigue tests on a servohydraulic tensile test machine operate at 60 Hz, requiring approximately 46 h to accumulate 10^7 cycles for each data point on the Goodman diagram. Therefore, significant amounts of time are required to characterize the fatigue properties of typical aerospace structural materials.

Furthermore, these uniaxial fatigue data are insufficient for assessing high-cycle fatigue (HCF) damage, which often occurs in components that are subjected to multiaxial loading over a wide range of cyclic frequencies including short-wavelength bending modes. Unfortunately, the current development of low-frequency multiaxial fatigue testing methods is experiencing a major drawback due to the high cost. In addition, high-frequency multiaxial fatigue methods currently do not exist. For example, existing biaxial fatigue machines have operating frequencies of less than 20 Hz, sometimes much less, on the order of 1 Hz. In turn, the development of a basic understanding of the effects of multiaxial loading on high-cycle fatigue is jeopardized. Hence, the current U.S. Air Force high-cycle fatigue research efforts have been significantly hindered due to lack of ability of obtaining a sufficiently large amount of multiaxial fatigue data for a meaningful fatigue strength assessment.

To address these concerns, a novel vibration-based fatigue testing concept was proposed [4,5] for assessing turbine engine material fatigue strength under multiple stress states found in the real engine hardware. The idea is to accumulate stress cycles on a base-excited plate specimen at high-frequency resonant modes such as the two-stripe mode (frequency range 1200–1600 Hz), frequencies that cannot be achieved in conventional fatigue test machines even under uniaxial stress. This goal can be achieved in three steps: (1) a topological design procedure, incorporating a finite element model (FEM), to characterize the shape of the specimens for ensuring the required stress state/pattern, (2) a vibration feedback empirical procedure for achieving the high-cycle fatigue experiments with variable-amplitude loading, and (3) a yielding procedure for achieving various uniaxial stress ratios.

2 Fatigue Specimen Topology Design

2.1 Uniaxial Fatigue Specimen. The first step to perform the vibration-based fatigue test is to search for a specimen geometry that possesses a mode shape within the operating frequency range of the shaker that produces a desirable stress field. A desirable stress field has a region of stress that is several times the magnitude of the maximum stress that occurs along the clamped edge to prevent fatigue initiating along the clamped edge. Specifically, for a uniaxial stress state in bending it is required that the maximum stress state occur along a free edge. A specimen geometry that was found using a topological design procedure developed by the authors [4–6] that produces such a stress state is a square plate, clamped and excited along one edge, vibrating in the mode shown in Fig. 2. In Fig. 2 the nodal degrees of freedom of the plate finite element model are constrained along the lower

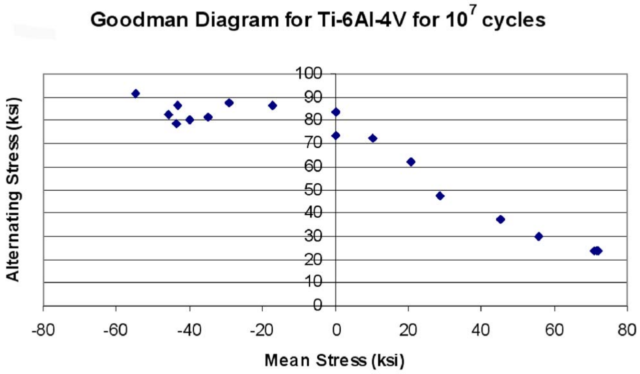


Fig. 1 Typical Goodman (or Haigh) diagram for Ti-6Al-4V for 10^7 cycles [3]

---

[1] Author to whom correspondence should be addressed.

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received October 2002, final manuscript received March 2003. Assoc. Editor: H. R. Simmons. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Atlanta, GA, June 16–19, 2003, Paper No. 2003-GT-38512.

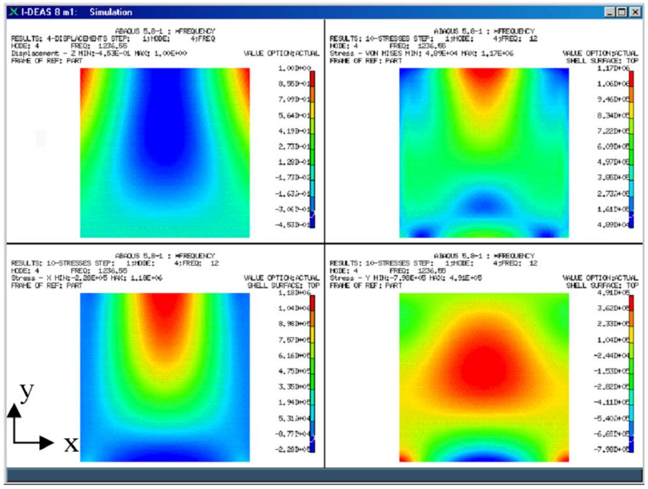


Fig. 2 FEM results for square steel plate geometry

edge of the model, resulting in a cantilever. The mode shape's out-of-plane vibratory motion is shown in the top left frame of Fig. 2, the corresponding von Mises stress is shown in the top right frame, the x-direction stress in the lower left frame, and the y-direction stress in the lower right frame. This mode shape is typically referred to as the two-stripe mode in reference to this mode shape's two nodal lines. These FEM results are for a 4.5-in. square 0.095-in. steel plate. As seen in the von Mises stress plot at the top right of Fig. 2, there exists a maximum stress in the plate that is located at the center of the free end. This location, the fatigue zone, is where fatigue failure is predicted to occur for this specimen geometry. Examination of the x and y direction stress plots in Fig. 2 indicates that the stress in the fatigue zone is uniaxial in the x direction. This must be the case since the fatigue zone lies on a free edge.

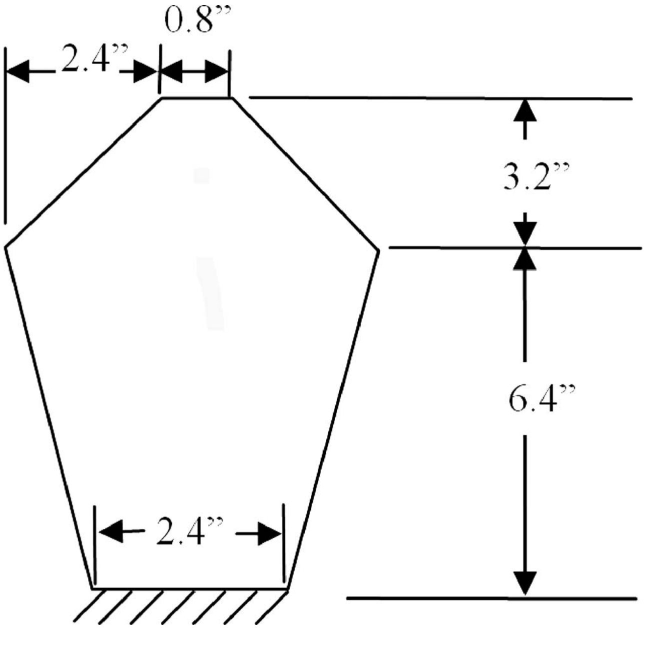


Fig. 3 Biaxial specimen Winger3 dimensions

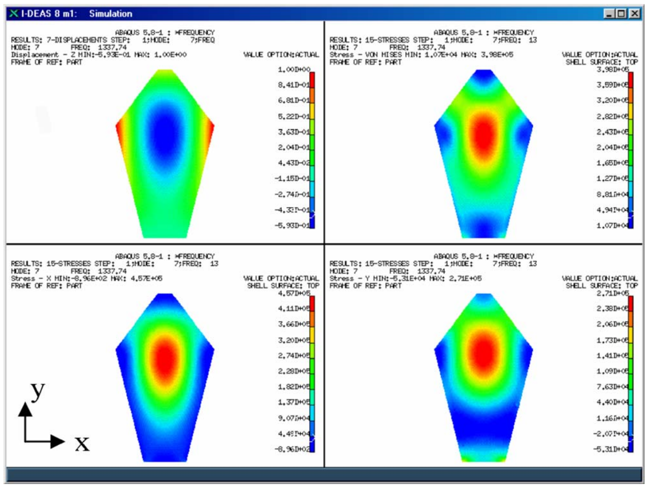


Fig. 4 FEM results for Alumwinger3 (0.125-in. thick 6061-T6)

2.2 Biaxial Fatigue Specimen. A cantilevered geometry named Winger3, shown in Fig. 3, was identified via the topological design process developed in Res. [4–6]. Due to the additional complexities of designing a biaxial specimen, this topological design process was performed manually. It is a matter of finding a plate geometry that has a mode shape with a biaxial stress state located away from the clamped edge while at the same time the von Mises stress in this biaxial stress region is at least a factor of 2 higher than the von Mises stress elsewhere in the plate to prevent fatigue failure along the clamped edge.

The FEM analysis of this geometry when constructed from 0.125-in. 6061-T6, called Alumwinger3, is shown in Fig. 4. The von Mises stress plot at the upper right of Fig. 4 shows that this geometry has stresses along the clamped edge approximately a factor of 2 less than that in the fatigue region, which meets our previously stated goal. An in-phase biaxial bending stress field with a ratio of $\sigma_Y/\sigma_X \cong 0.59$ is developed in the upper center of the plate as shown by the x and y stress plots in the lower left and right, respectively, of Fig. 4.

3 Empirical Procedure

Systematic stress control fatigue tests were conducted using the Unholtz-Dickie 6,000-lb electrodynamic shaker located in the Turbine Engine Fatigue Facility (TEFF) of the Air Force Research Laboratory (AFRL) in the Propulsion Directorate (AFRL/PR) at Wright-Patterson Air Force Base (WPAFB). The test specimen is mounted cantilevered to the shaker head as shown in Fig. 5. The FEM results, as shown in Figs. 2 and 4, were used to locate the instrumentation on the plate specimens.

The measurement instruments used consist of accelerometers, a laser vibrometer capable of measuring the velocity or displacement of a point on the specimen, and strain gages. In most cases, tests are performed using the laser vibrometer for measurement as this provides the only nonintrusive technique of the available instrumentation. However, the velocity signal from the laser must first be calibrated to the corresponding strain in the fatigue region of the plate. This is done by running a calibration test with the plate specimen instrumented with a strain gage placed in the expected fatigue region.

Data were taken from both the strain gage and from the laser vibrometer while the specimen was in resonance at various shaker power settings, or in other words, for various strain levels in the specimen, which allows for the construction of a strain–laser calibration curve. Note the laser must be positioned at a location that will not produce an overload or clipping of its amplifier when exciting the plate at maximum test levels. Once a relationship

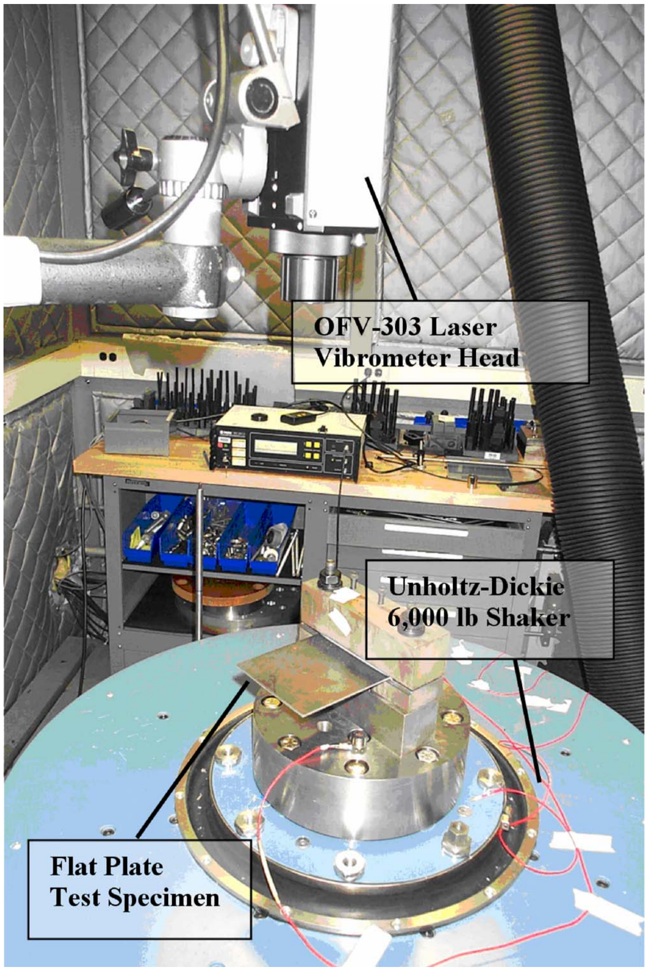


Fig. 5 Experimental setup

between the strain in the fatigue region and the velocity/displacement of the calibrated laser position is established, a fatigue test may be conducted.

The step test method developed by Maxwell and Nicholas [7] may be used to determine material endurance limits for a particular stress ratio R. The stress ratio is defined as $R=\sigma_{\min}/\sigma_{\max}$, where $\sigma_{\min}$ and $\sigma_{\max}$ are the minimum and maximum stresses in the fatigue region over a vibratory cycle. For the present situation, the stress state in the fatigue zone alternates between $+\sigma$ and $-\sigma$, or in other words the stress state is fully reversed and $R=-1$. This step test technique uses a single specimen to generate an endurance limit from multiple incremental loading steps using the following equation:

$$\sigma_A=\sigma_{\mathrm{pr}}+\frac{N_f}{10^m}(\sigma_f-\sigma_{\mathrm{pr}}), \tag{1}$$

where σ_A is the alternating stress to be plotted on a Goodman diagram, σ_{pr} is the alternating stress level of the step previous to failure, N_f is the number of cycles to failure in the final step, σ_f is the alternating stress level of the final (failure) step, and 10^m gives the number of cycles you are testing to.

The test is begun at an alternating stress level below that of where fatigue failure is expected, for the number of cycles of interest and in consideration of the particular test specimen material. After completing the specified number of vibratory stress cycles without a failure the stress level is increased. The test is then performed at this new stress level until either failure or the specified number of cycles is reached. This process is repeated until failure occurs. Then using the above equation the endurance limit of the material at the desired number of cycles can be determined. Essentially, this method is intended to rapidly generate data for the construction of a Goodman diagram.

The fatigue limit is defined in typical tests, performed using uniaxial machines, as the point at which the fatigue crack propagates through the specimen and failure occurs. Computations have shown that the fatigue crack propagation life is only a small fraction of total life when testing at stress levels near the HCF limit [8]. In the case of vibration-based fatigue testing, a definitive phenomenon such as abrupt failure does not exist. Therefore, in this technique the fatigue limit is defined at the instance corresponding to a sudden change in the dynamic response of the plate associated with the initial stage of fatigue crack development. The development of a crack in the specimen changes its stiffness and hence its resonant frequency. The initiation or development of the fatigue crack in the uniaxial and biaxial specimens was observed by the onset of a rapid decrease in the measured velocity for a given shaker driving frequency and amplitude and it is at this point that the fatigue limit stress was determined.

During the step method fatigue test, whenever the stress level dropped approximately 5% below the desired stress level, the shaker driving frequency and amplitude were adjusted to maintain the desired resonant mode and amplitude of the plate at the desired stress level. As soon as a crack was identified the shaker amplitude was left unchanged and the driving frequency was reduced as necessary until the plate would no longer excite to the desired stress level in order to develop the crack to a visible length. The observed fatigue cracks produced using this techinique were on the order of 1–2 mm in the case of steel specimens and 0.1 mm in the case of Ti-6Al-4V.

4 Results and Discussion

The materials tested using the vibration-based test method were steel, 6061-T6 aluminum, and the titanium alloy Ti-6Al-4V. These materials were not characterized and were used only to demonstrate the vibration-based test methodology. The steel and 6061-T6 aluminum were chosen as these materials were readily available within the TEFF Laboratory. The Ti-6Al-4V material was purchased as a sheet from a commercial supplier and was manufactured to ASTM B265 specifications. The fatigue results obtained with these materials should be considered only as demonstrations of the test methodology.

4.1 Uniaxial Fatigue Tests. Using the step test method of Maxwell and Nicholas described previously, three data points were obtained for the fatigue limit stress of unnotched mild steel specimens, in this case at 10^7 cycles. The natural frequency of the two-stripe mode for the steel specimen used was approximately 1200 Hz. Therefore, a fatigue test step lasting 10^7 cycles can be performed in roughly 2 h 18 min compared to a similar test at 60 Hz, which would take over 46 h. This represents approximately a 20× decrease in test time. The experimental fatigue results for 4.5-in.-square (clamped dimensions), 0.095-in.-thick (measured

Table 1 Steel experimental results for 10^7 cycles

Plate no.	σ_{pr} (ksi)	σ_{f} (ksi)	σ_{A} (ksi)
OSU#7	30	35	31.6
OSU#8	35	40	36.2
OSU#9	35	40	38.3

Table 2 6061-T6 experimental results for 10^6 cycles

Plate no.	σ_{pr} (ksi)	σ_{f} (ksi)	σ_{A} (ksi)
Al6	25	30	25.2
Al7	25	30	27.4

Table 3 Ti-6Al-4V experimental results for 10^6 cycles

Plate no.	σ_{pr} (ksi)	σ_f (ksi)	σ_A (ksi)
1	75	80	77.5

0.093-in.) unnotched steel uniaxial specimens are given in Table 1; note the data also show the scatter that is typical for fully reversed stress.

The fatigue results for unnotched 4.5-in. square, 0.125-in.-thick (measured 0.122-in.) specimens of the aluminum alloy, 6061-T6, were determined using the vibration-based fatigue test method for 10^6 cycles. These results are shown in Table 2. It should be noted for this and the following test cases that testing was performed to 10^6 cycles instead of 10^7 cycles due to laboratory noise concerns as well as time constraints due to equipment availability. The specimen geometry used had a two-stripe mode natural frequency of approximately 1600 Hz resulting in a fatigue test step lasting roughly 10 min compared to over 4.6 h to accumulate 10^6 cycles for a test machine operating at 60 Hz.

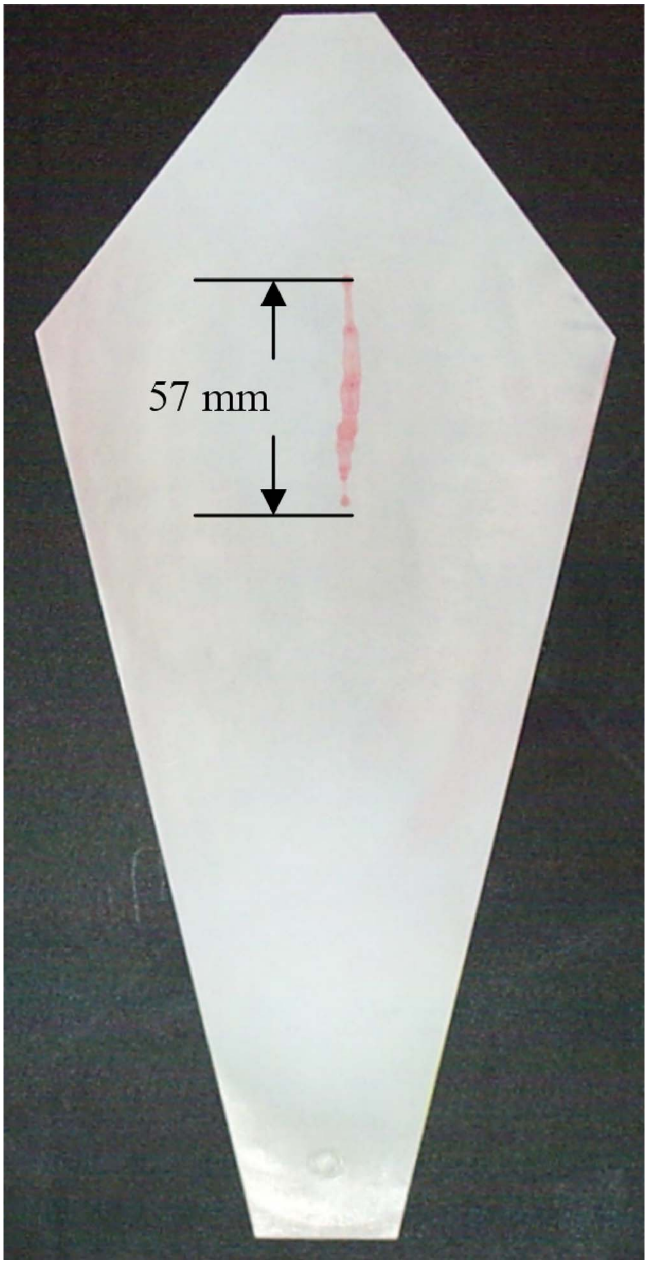


Fig. 6 Photograph of cracked biaxial specimen Alumwinger3

Table 4 Biaxial fatigue results for 6061-T6 for 10^6 cycles

Plate no.	σ_{pr} (ksi)	σ_f (ksi)	$(\sigma_A)_{equiv}$ (ksi)
4	15	20	18.8
5	15	20	19.4

Table 5 Comparison of uniaxial and biaxial 6061-T6 data for 10^6 cycles

Average uniaxial fatigue strength (ksi)	26.3
Average biaxial von Mises equivalent strength (ksi)	19.1

The fatigue result at 10^6 cycles for a 4.5-in. square, 0.125-in.-thick (measured 0.117-in.) unnotched specimen of the popular aerospace titanium alloy, Ti-6Al-4V, is shown in Table 3. The specimen geometry used had a two-stripe mode natural frequency of approximately 1500 Hz, resulting in a fatigue test step lasting roughly 11 min for 10^6 cycles compared to over 4.6 h for a test machine operating at 60 Hz.

4.2 Biaxial Fatigue Tests. Biaxial fatigue experiments were conducted on the Winger3 specimen made of a 6061-T6, 0.125-in.-thick aluminum sheet, so-called Alumwinger3. Figure 6 shows a fatigue crack that was developed and propagated significantly in an Alumwinger3 specimen. Note that the crack propagation direction is perpendicular to the dominant stress field, in this case the x-direction stress dominates the y-direction stress as shown in Fig. 4.

Given the harmonic nature of the specimen resonant response to the sinusoidal excitation of an electrodynamic shaker, the x to y stress ratio is proportional and, like in the case of the uniaxial specimen, fully reversed. To allow for comparison with uniaxial stress fatigue results, the biaxial fatigue strength results are given in the form of the von Mises equivalent stress

$$\sigma_{equiv}=\frac{1}{\sqrt{2}}\sqrt{(\sigma_1-\sigma_2)^2+(\sigma_2-\sigma_3)^2+(\sigma_3-\sigma_1)^2}, \qquad (2)$$

where σ_1, σ_2, and σ_3 are the principal stresses. In the biaxial case in the fatigue zone of the Alumwinger3 geometry $\sigma_3=0$ and σ_1 and σ_2 are nonzero and experimentally determined from laser data calibrated to data from a delta rosette strain gage applied at the fatigue zone. The fatigue results at 10^6 cycles for two 6061-T6 biaxial test specimens are given in Table 4. The averaged von Mises equivalent stress biaxial fatigue results are compared to the previous uniaxial fatigue results average for 6061-T6 and are presented in Table 5. Table 5 shows that the von Mises equivalent fatigue strength results for a biaxial stress state are not consistent and significantly less than that obtained for uniaxial stress.

5 Small Fatigue Crack Identification

The biaxial specimen frequency response showed significant nonlinear hardening that can be made use of for crack identification. The development of a fatigue crack in this case was observed by a relatively rapid increase in the measured plate velocity obtained with the laser vibrometer quickly followed by a very discrete drop in response as the resonant frequency shifted enough relative to the driving frequency that the response fell off of the nonlinear hardening peak. This phenomenon made possible the identification of a crack in the early initiation stage from this vibration-based test method that cannot be achieved using other existing test methods. Figure 7 shows an approximately 0.4-mm-long series of microcracks that were developed and identified in a biaxial specimen.

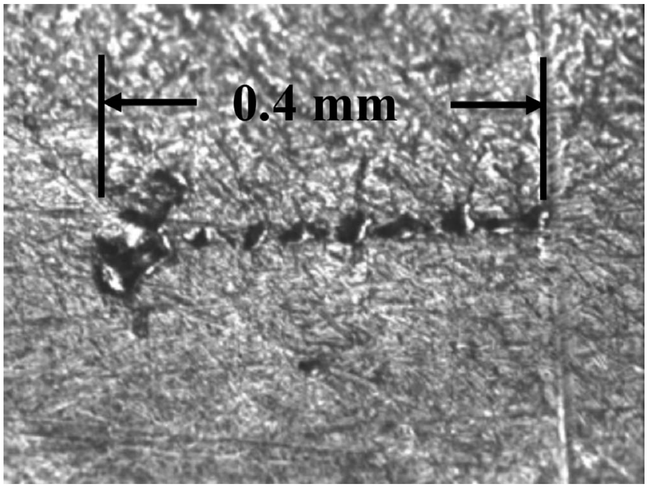
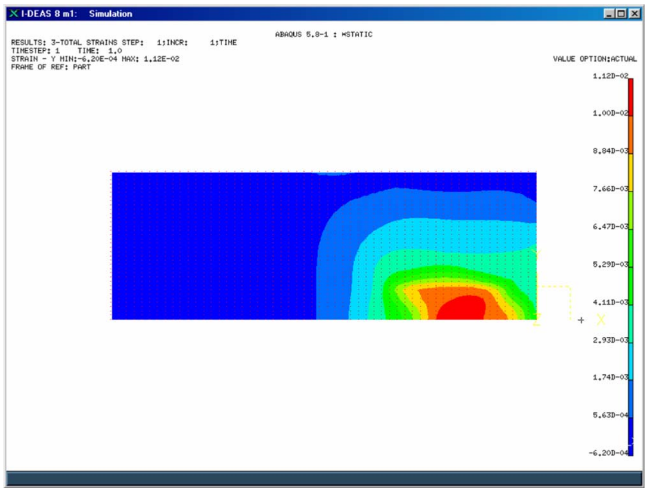


Fig. 7 Microphotographs of microcracks in a 6061-T6 biaxial specimen



```
                                        ABAQUS 5.8-1 : *STATIC
RESULTS: 3-TOTAL STRAINS STEP:   1;INCR:     1;TIME
TIMESTEP: 1    TIME:  1.0
STRAIN - Y MIN:-6.20E-04 MAX: 1.12E-02                                            VALUE OPTION:ACTUAL
FRAME OF REF: PART
```

Fig. 9 Maximum load *y*-direction total strain

6 Variation of Stress Ratio

To enable the construction of the Goodman diagram for fatigue design, a methodology was developed to produce fatigue strength data with variation of stress ratios using the vibration-based testing procedure. In typical uniaxial test machines this is a straightforward task. With the flexibility of servohydraulics it is a simple matter to adjust the mean stress level by commanding the test machine to apply a static load in addition to the alternating load. For the situation of simple vibratory motion, as on the TEFF Laboratory electrodynamic shakers, the stress is fully reversed and the mean stress is zero.

In this investigation, the idea of using a yielding procedure to generate residual stresses at the desired locations in a plate specimen was explored. The idea is presented as follows. Typically, yielding tests are conducted on so called dogbone tensile test specimens or coupons. In this case the yielding takes place throughout the entire cross section and very little residual stresses remain after the specimen is unloaded. However, if the yielding does not occur throughout the entire cross section but rather only locally, the material surrounding the localized plastic deformation will try to elastically spring back to establish equilibrium, thus creating residual elastic strain and therefore residual stresses. Therefore, by controlling the external loading conditions (magnitude, location, and loading area), desired residual stresses in the fatigue zone of the uniaxial plate specimen can be achieved to enable vibration-based fatigue testing at various *R* ratios.

The TEFF Laboratory's MTS Systems Corporation 22,000 lb uniaxial load frame was used for the residual stress producing

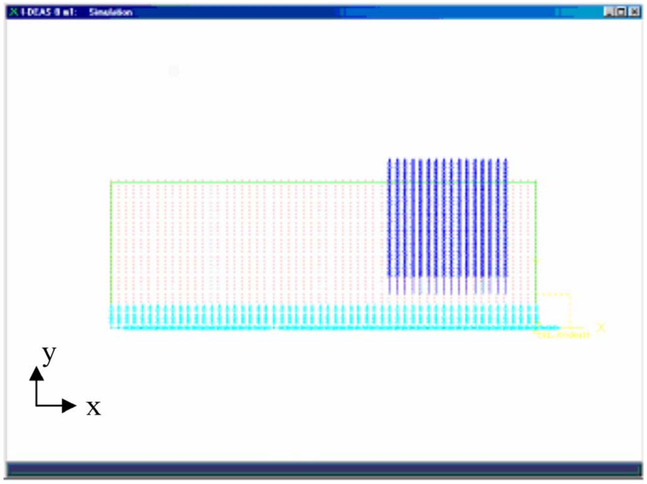
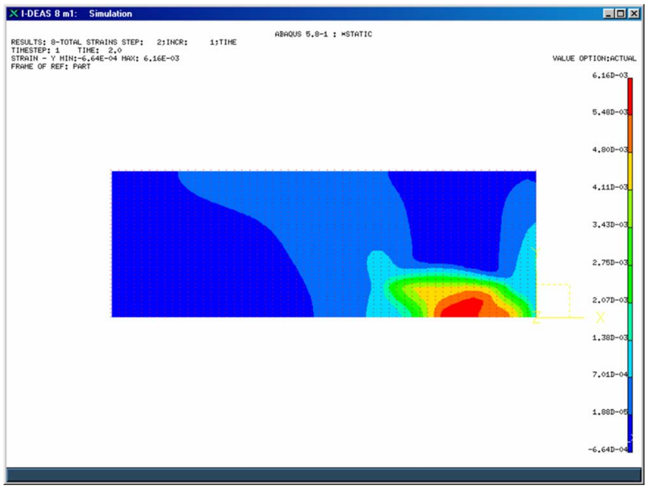
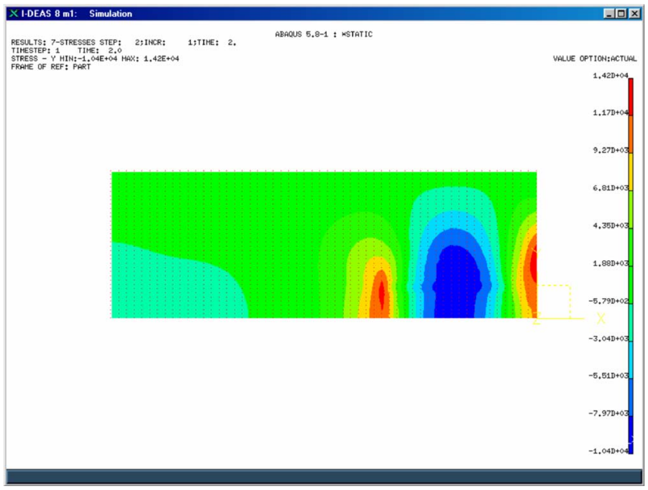


Fig. 8 Loading location



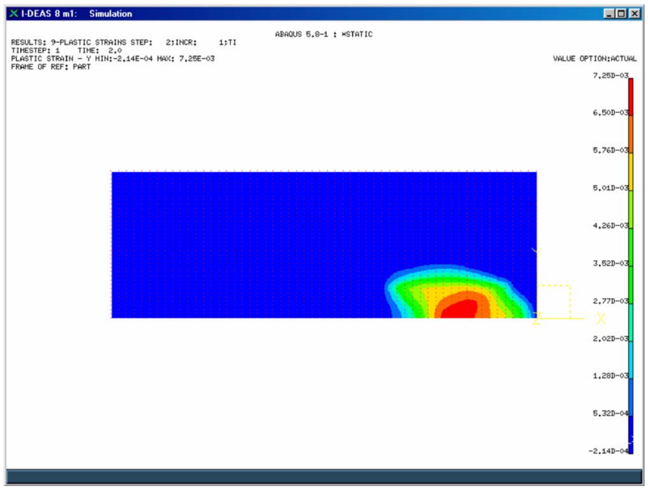
```
                                        ABAQUS 5.8-1 : *STATIC
RESULTS: 8-TOTAL STRAINS STEP:   2;INCR:     1;TIME
TIMESTEP: 1    TIME:  2.0
STRAIN - Y MIN:-6.64E-04 MAX: 6.16E-03                                            VALUE OPTION:ACTUAL
FRAME OF REF: PART
```

Fig. 10 Postyield *y*-direction total strain



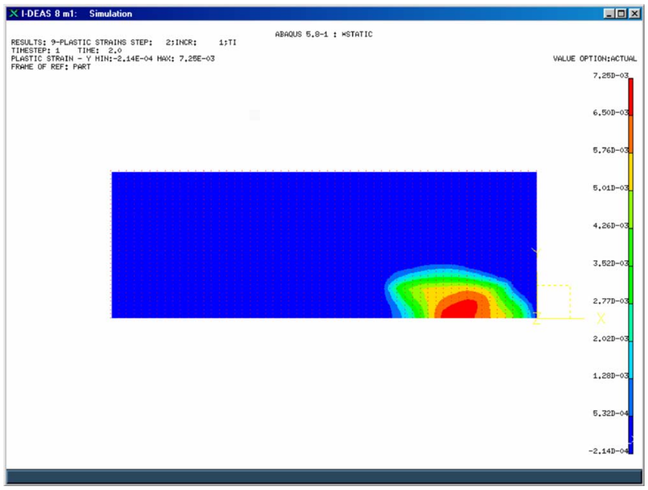
```
                                        ABAQUS 5.8-1 : *STATIC
RESULTS: 7-STRESSES STEP:   2;INCR:     1;TIME:  2.
TIMESTEP: 1    TIME:  2.0
STRESS - Y MIN:-1.04E+04 MAX: 1.42E+04                                            VALUE OPTION:ACTUAL
FRAME OF REF: PART
```

Fig. 11 Postyield *y*-direction stress



```
                                   ABAQUS 5.8-1 : *STATIC
RESULTS: 9-PLASTIC STRAINS STEP:   2;INCR:     1;TI
TIMESTEP: 1    TIME:  2.0
PLASTIC STRAIN - Y MIN:-2.14E-04 MAX: 7.25E-03                          VALUE OPTION:ACTUAL
FRAME OF REF: PART
```

Fig. 12 Postyield *y*-direction elastic strain



```
                                   ABAQUS 5.8-1 : *STATIC
RESULTS: 9-PLASTIC STRAINS STEP:   2;INCR:     1;TI
TIMESTEP: 1    TIME:  2.0
PLASTIC STRAIN - Y MIN:-2.14E-04 MAX: 7.25E-03                          VALUE OPTION:ACTUAL
FRAME OF REF: PART
```

Fig. 13 Postyield *y*-direction plastic strain

Table 6 Experimental load–strain results for the 6061-T6 specimens

	Maximum load data			Zero-load residual strain	
Plate no.	Load (lb)	$\varepsilon_y^{\text{total}}$ (10^6) gage 1	$\varepsilon_y^{\text{total}}$ (10^6) gage 2	$\varepsilon_y^{\text{total}}$ (10^6) gage 1	$\varepsilon_y^{\text{total}}$ (10^6) gage 2
1	14500	4271	6925	1240	1710
2	16000	8099	11195	4285	5594
3	14500	gage failed	6824	gage failed	1793
4	16000	8053	11063	4284	5480

yielding procedure of uniaxial test specimens. This machine is equipped with MTS 3000-psi hydraulic wedge grips, which can utilize 1.75-in.-wide diamond face wedges. The design of the hydraulic wedge grip limited the useful depth of the uniaxial test specimen that could be gripped to approximately 1.75 in. Elastic–plastic FEM analyses were performed to identify the proper location to place the grips on the uniaxial test specimen.

The uniaxial plate specimen was modeled using I-DEAS for pre- and postprocessing and ABAQUS was used for the elastic–plastic analysis. The 6061-T6 uniaxial test specimen dimensions were 6.5 in.×4.5 in.×0.122 in., nominally. An elastic–plastic analysis using ABAQUS was performed using a 55×35×4 mesh, 55 elements across the 6.5-in. span, 35 elements across across the 2.25-in. half chord, and 4 elements through the 0.122-in. thickness. The plate was modeled using half-plane symmetry. The material properties used for the 6061-T6 were $E=1.0E+07$ psi, $\nu=0.33$, $\sigma_{\text{yield}}=39900$ psi, and the material behavior was assumed to be elastic, perfectly plastic. Representative grip loads were applied in the *y* direction, perpendicular to the long axis of the plate. Figure 8 shows the loading location that produced the desired postyield residual *y*-direction tensile stress in the fatigue zone. The *y*-direction total strains are shown for both the maximum loading and the unloaded or post-yielding case in Figs. 9 and 10, respectively. Figure 11 shows the postyield *y*-direction residual stress field. The postyield *y*-direction residual elastic strain and the *y*-direction plastic strain are shown in Figures 12 and 13, respectively.

The total strain field results from the FEM analysis were used to locate the instrumentation on the plate. Two Micro-Measurements CEA-05-062UW-350 strain gages were placed as shown in Fig. 14 along the spanwise centerline of the 6061-T6 plate at 0.96 in. from the right edge and at 0.125 in. from the right, which is the closest to the edge that the strain gage dimensions will allow.

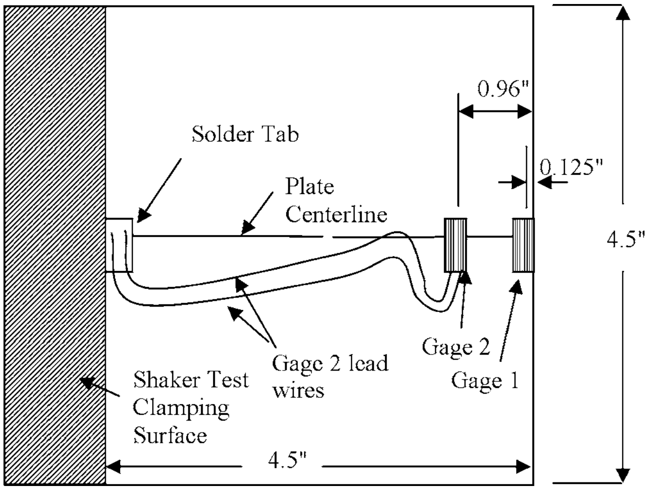


Fig. 14 Diagram of strain gage placement

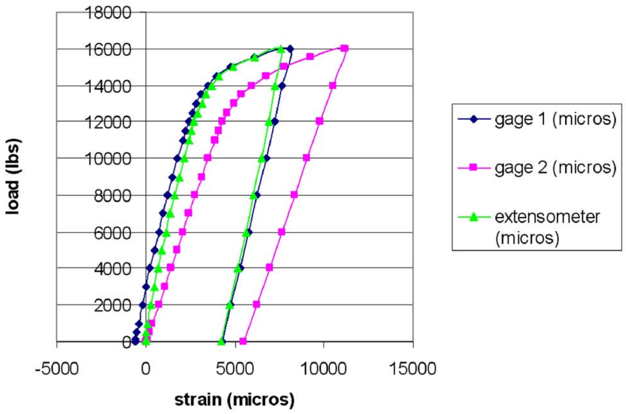


Fig. 15 Load–strain history for 6061-T6 specimen 2

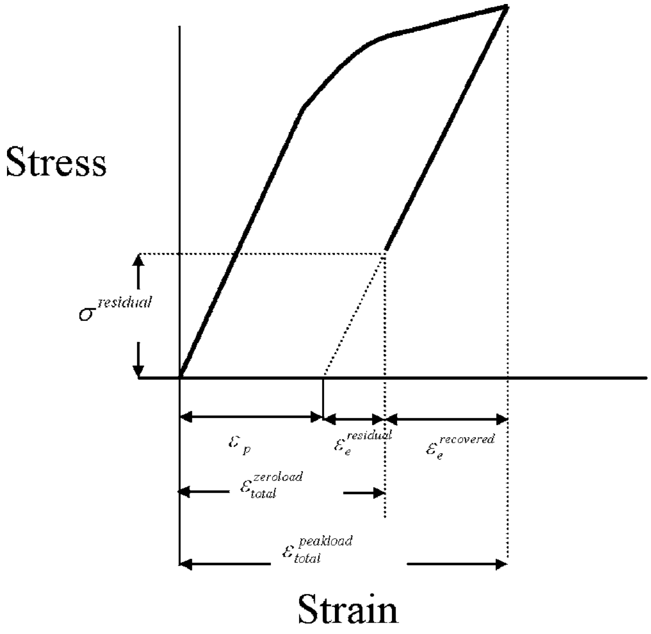


Fig. 16 Total, elastic, plastic strains and residual stress

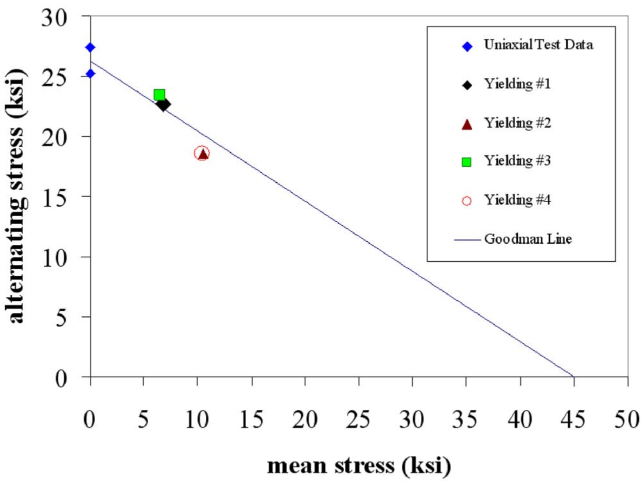


Fig. 17 Goodman diagram for 6061-T6 for 10^6 cycles from vibration-based test data

The 0.125-in. location of gage 1 is centered in the fatigue zone that is created by the two-stripe mode and the 0.96-in. location is in the region of maximum total strain at the maximum load during the yielding test. An MTS extensometer was also attached to the plate edge near the gage 1 location to monitor the strain and to verify the strain gage readings.

Using the maximum FEM total strain at the gage 2 location as a target, the yielding procedure was performed on the MTS machine and loading was adjusted to closely achieve the FEM strain level. The FEM analysis was then repeated, adjusting the loading in the FEM until the experimentally measured total strain at the gage 2 location was matched.

Four plates were yielded using the yielding procedure described above. The readings at maximum load and at zero load are given in Table 6.

Figure 15 shows the load versus strain history for 6061-T6, yielding specimen 2. The extensometer reading agrees well with the readings from gage 1 and hence only strain readings from the two strain gages were considered.

In the situation of local yielding, residual stresses remain and the stress does not necessarily return to zero but follows a similar stress–strain curve to the solid stress-strain path shown in Fig. 16. Of course, the strain gages and extensometer measure the total strain. What is of importance for subsequent vibration-based fatigue testing of the yielded specimens is the value of the residual elastic strain and hence the residual stress. The only method readily available to us for separating the elastic strain from the measured total strain is our previously described elastic–plastic FEM analysis. The FEM predictions of the residual total strains at the locations of gages 1 and 2 are given in Table 7. Table 8 lists the FEM predicted residual tensile stresses at the location of gage 1 which is the fatigue zone for the uniaxial specimen described in Sec. 2.1.

With our best estimate made for the residual stress in the fatigue zone, vibration-based fatigue tests were conducted of the yielded 6061-T6 plates. The vibration-based fatigue procedure used was that described in Sec. 3. Table 9 lists the alternating fatigue stress results for 10^6 cycles for the four yielded 6061-T6 plates.

The uniaxial data for the fully reversed case presented in Table 2 are compared with the data for the four yielded plates in Fig. 17. The Goodman line in Fig. 17 represents the average of the fully reversed experimental data connected through the ultimate tensile strength of 6061-T6 specimens. Figure 17 shows that the technique of using a local yielding procedure to produce a residual stress field, enabling vibration-based testing at various R ratios, has promise. Further investigation is required to determine if the yielding procedure has affected the fatigue life of the 6061-T6 material.

Table 7 FEM load–strain results for the 6061-T6 specimens

	Maximum load data			Zero-load residual strain	
Plate no.	Load (lb)	ε_y^{total} (10^6) gage 1	ε_y^{total} (10^6) gage 2	ε_y^{total} (10^6) gage 1	ε_y^{total} (10^6) gage 2
1	12759	3180	6925	260	2525
2	13875	5970	11195	2700	5600
3	12683	3000	6824	240	2380
4	13861	5500	11063	2720	5530

Table 8 FEM residual stress estimates for the 6061-T6 specimens

Plate no.	$\sigma_y^{residual}$ (ksi)
1	6.75
2	10.5
3	6.50
4	10.4

Table 9 Alternating fatigue results for 6061-T6 for 10^6 cycles from vibration-based test data

Plate no.	σ_{pr} (ksi)	σ_f (ksi)	σ_A (ksi)
1	20	25	22.7
2	17.5	20	18.6
3	22.5	25	23.4
4	17.5	20	18.6

7 Conclusions

The test method described in this paper will allow aerospace material fatigue properties to be characterized at greatly reduced savings in both time and cost. Fatigue testing to 10^7 cycles can be accomplished in only a few hours compared to the tens of hours required of typical/conventional fatigue test machines. This vibration-based method is capable, through proper specimen design, of generating both uniaxial and biaxial data. A promising technique was developed to produce residual stress enabling vibration-based fatigue testing at various stress ratios. The performance of the technique for achieving various stress ratios was demonstrated with experimental results from 6061-T6 aluminum plate specimens. The resulting fatigue data at various stress ratios have been used to construct the important Goodman diagram for fatigue design.

Acknowledgment

The authors would like to thank the Dayton Area Graduate Studies Institute (DAGSI) and the Air Force Research Laboratories, specifically the Materials Laboratory and the TEFF Laboratory, for their financial support, facility and equipment access, and encouragement of this research.

References

[1] Goodman, J., 1899, *Mechanics Applied to Engineering*, Longmans, Green, and Co., London.

[2] Nicholas, T., and Zuiker, J. R., 1996, "On the Use of the Goodman Diagram for High Cycle Fatigue Design," Int. J. Fract., **80**, pp. 219–235.

[3] Nicholas, T., and Maxwell, D. C., 2002, "Mean Stress Effects on the High Cycle Fatigue Limit Stress in Ti-6Al-4V," *Fatigue and Fracture Mechanics*: Vol. 33, ASTM STP 1417, Reuter, W. G., and Piascik, R. S., eds., American Society for Testing and Materials, West Conshohocken, PA, pp. 476–492.

[4] Shen, M.-H. H., Seidt, J., George, T., Cross, C., Whaley, P. W., and Nicholas, T., 2001, "Development of A Novel Method for Evaluating Material Behavior under Turbine Engine Operating Conditions, Part I: Design of Accelerated HCF Testing Procedures," 6th National Turbine Engine High Cycle Fatigue Conference, Jacksonville, FL, March 5-8.

[5] Shen, M.-H. H., George, T., Seidt, J., Nicholas, T., and Cross, C., 2001, "Development of A Novel Method for Evaluating Material Behavior under Turbine Engine Operating Conditions, Part II: An Empirical Vibration-Based Fatigue Assessment Framework," 6th National Turbine Engine High Cycle Fatigue Conference, Jacksonville, FL, March 5-8.

[6] Seidt, J., 2001, "Development of a Novel Vibration Based High Cycle Fatigue Test Method," M.S. thesis, The Ohio State University, Columbus, OH.

[7] Maxwell, D. C., and Nicholas, T., 1999, "A Rapid Method for Generation of a Haigh Diagram for High Cycle Fatigue," *Fatigue and Fracture Mechanics*: Vol. 29, ASTM STP 1321, T. L. Panontin, and S. D. Sheppard, eds., American Society for Testing and Materials, West Conshohocken, PA, pp. 626–641.

[8] Morrissey, R. J., Golden, P., and Nicholas, T., "The Effects of Stress Transients on the HCF Endurance Limit," Int. J. Fatigue, **25**(9–11), pp. 1125–1133.

Case Studies of Fatigue Life Improvement Using Low Plasticity Burnishing in Gas Turbine Engine Applications

Paul S. Prevéy
Lambda Research,
5521 Fair Lane,
Cincinnati, OH 45227
e-mail: pprevey@lambda-research.com

Ravi A. Ravindranath
NAVAIR,
22195 Elmer Road, Building 106, Room 202-G,
Patuxent River, MD 10670-1534
e-mail: ravindranara@navair.navy.mil

Michael Shepard
Wright Patterson AFB,
2230 Tenth Street, Ste. 1,
Wright Patterson AFB, OH 45433-7817
e-mail: michael.Shepard@wpafb.af.mil

Timothy Gabb
NASA Glenn Research Center,
21000 Brookpark Road,
Building 49, Room 231,
Cleveland, OH 44135-3191
e-mail: timothy.gabb@lerc.nasa.gov

Surface enhancement technologies such as shot peening, laser shock peening, and low plasticity burnishing (LPB) can provide substantial fatigue life improvement. However, to be effective, the compressive residual stresses that increase fatigue strength must be retained in service. For successful integration into turbine design, the process must be affordable and compatible with the manufacturing environment. LPB provides thermally stable compression of comparable magnitude and even greater depth than other methods, and can be performed in conventional machine shop environments on CNC machine tools. LPB provides a means to extend the fatigue lives of both new and legacy aircraft engines and ground-based turbines. Improving fatigue performance by introducing deep stable layers of compressive residual stress avoids the generally cost prohibitive alternative of modifying either material or design. The x-ray diffraction based background studies of thermal and mechanical stability of surface enhancement techniques are briefly reviewed, demonstrating the importance of minimizing cold work. The LPB process, tooling, and control systems are described. An overview of current research programs conducted for engine OEMs and the military to apply LPB to a variety of engine and aging aircraft components are presented. Fatigue performance and residual stress data developed to date for several case studies are presented including the following. (1) The effect of LPB on the fatigue performance of the nickel based super alloy IN718, showing the fatigue benefit of thermal stability at engine temperatures. (2) An order of magnitude improvement in damage tolerance of LPB processed Ti-6-4 fan blade leading edges. (3) Elimination of the fretting fatigue debit for Ti-6-4 with prior LPB. (4) Corrosion fatigue mitigation with LPB in Carpenter 450 steel. (5) Damage tolerance improvement in 17-4 PH steel. Where appropriate, the performance of LPB is compared to conventional shot peening after exposure to engine operating temperatures. [DOI: 10.1115/1.1807414]

Introduction

LPB is a new method of surface enhancement [1–4] that provides deep stable surface compressive residual stresses with little cold work for improved fatigue, fretting fatigue, and stress corrosion performance even at elevated temperatures where compression from shot peening relaxes [5]. LPB surface treatment is applied using conventional multi-axis CNC machine tools for unprecedented control of the residual stress distribution developed through modification of the pressure, feed, and tool characteristics. The resulting deep layer of compressive residual stress has been shown to improve high cycle fatigue (HCF) and low cycle fatigue (LCF) performance and foreign object damage (FOD) tolerance [6,7]. Achieving deep compression with low cold work (<5%) reduces relaxation of the protective compressive layer either thermally during exposures at service temperatures, or mechanically due to overload or impact, as in FOD. Both thermal and mechanical relaxation occur in the cold worked surfaces produced by conventional shot peening of titanium [7] and nickel alloys [5,6].

The LPB tooling and process are described and the protocol for use of LPB in fatigue-critical design is presented. LPB has been applied to titanium, iron, and nickel based aero-turbine engine alloys to improve damage tolerance in compressor sections by an order of magnitude and to improve high and low cycle fatigue performance in hot turbine sections. In addition to turbine engine components, this novel technology has been shown to have potential applications in aging aircraft structures, nuclear waste-material containers, biomedical implants and welded joints. This paper attempts to present a brief overview of the fatigue and damage tolerance improvement achievable with LPB in aero engine components.

The Low Plasticity Burnishing (LPB) Process. LPB is a method of CNC controlled burnishing designed to produce a deep layer of highly compressive residual stress with a minimum amount of cold working, or plastic deformation. Residual compressive stresses approaching the material yield strength are developed using a series of passes of a freely rotating ball tool producing an accumulated plastic strain, or cold work, level of less than 3% to 5%. In contrast, the multiple random shot impacts of conventional shot peening produce cold work levels ranging from 20% to over 100%, leaving a severely deformed surface layer with a high dislocation density that adversely affects the thermal and mechanical stability of the compressive layer [5,7].

Unlike LPB, conventional roller and ball burnishing utilize a hard wheel tool or fixed lubricated ball pressed into the surface of an axisymmetric work piece with sufficient force to deform the near surface layers, usually in a lathe. Burnishing is performed with multiple passes, usually under increasing load, to improve surface finish and to deliberately cold work the surface. Roller and ball burnishing have been studied in Russia and Japan, and were applied most extensively in the USSR in the 1970s. Various burnishing methods are used, particularly in Eastern Europe to im-

---

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received October 2002; final manuscript received March 2003. Assoc. Editor: H. R. Simmons. Paper Presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Atlanta, GA, June 16–19, 2003, Paper No. 2003-GT-38922.

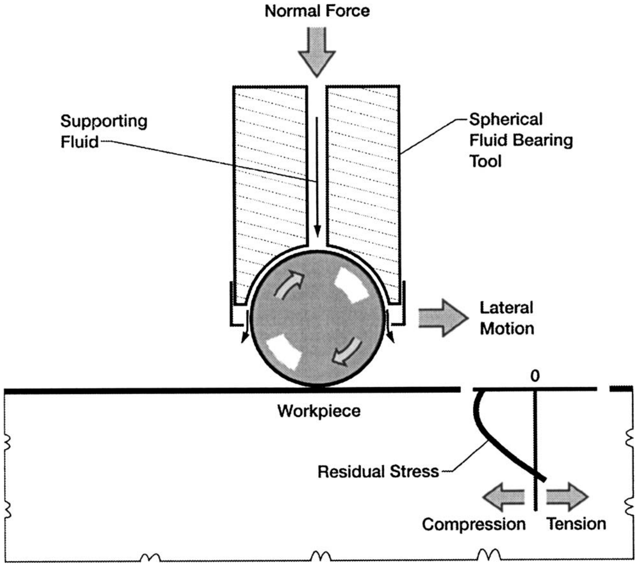


Fig. 1 LPB schematic

prove fatigue life. Improvements in HCF, corrosion-fatigue, and stress corrosion cracking are documented, with fatigue strength enhancement attributed to improved finish, the development of a compressive surface layer, and the increased yield strength of the cold worked surface [8–13]. Optimum forces and rolling parameters were established to minimize roughness and/or maximize surface hardening [14–17]. Analytical models to predict the residual stresses have been attempted in England [18] and France [19]. "Deep rolling" employs either roller tools or a partially hydrostatically supported burnishing ball, but differs from LPB in the use of higher loads and cold work levels, tool design, and control. X-ray diffraction line broadening and micro-hardness reveal that deep rolling produces even more cold work than shot peening [20–22].

The basic LPB tool is comprised of a ball that is supported in a spherical hydrostatic bearing as shown in Fig. 1. The tool can be held in any CNC lathe or mill. The machine tool coolant is used to pressurize the bearing with a continuous flow of fluid to support the ball. The ball does not contact the bearing seat, even under load. The ball is loaded normal to the surface of a component with a hydraulic cylinder that is in the body of the tool. LPB can be performed in conjunction with chip forming machining operations in the same CNC machine tool.

The ball rolls across the surface of a component in a pattern defined in the CNC code, as in any machining operation. The tool path and normal pressure applied are designed to create a distribution of compressive residual stress. The form of the distribution is designed to counter applied stresses and optimize fatigue or stress corrosion performance. Since there is no shear being applied to the ball, it is free to roll in any direction. As the ball rolls over the component, the pressure from the ball causes plastic deformation to occur in the surface of the material under the ball. Since the bulk of the material constrains the deformed area, the deformed zone is left in compression after the ball passes. No material is removed during the process. The surface is permanently displaced inward by only a few ten-thousandths of an inch (0.002–0.015 mm) (0.0001–0.0006 in.). LPB smoothes surface asperities leaving an improved surface finish that can be better than 5 μin., RA. Dual sided LPB processing of a 17-4 PH compressor blade with caliper tooling and a larger finished Ti-6-4 blade are shown in Figs. 2(a) and 2(b).

X-ray Diffraction, Residual Stress and Cold Work Determination. XRD residual stress measurements were made employing a $\sin^2 \psi$ technique and the diffraction of the appropriate radiation. It was first verified that the lattice spacing was a linear function of $\sin^2 \psi$ as required for the plane stress linear elastic residual stress model [23–26].

Measurements were made as a function of depth by incremental electropolishing to the final depth shown. The residual stress measurements were corrected for both the penetration of the radiation into the subsurface stress gradient [27] and for stress relaxation caused by layer removal [28].

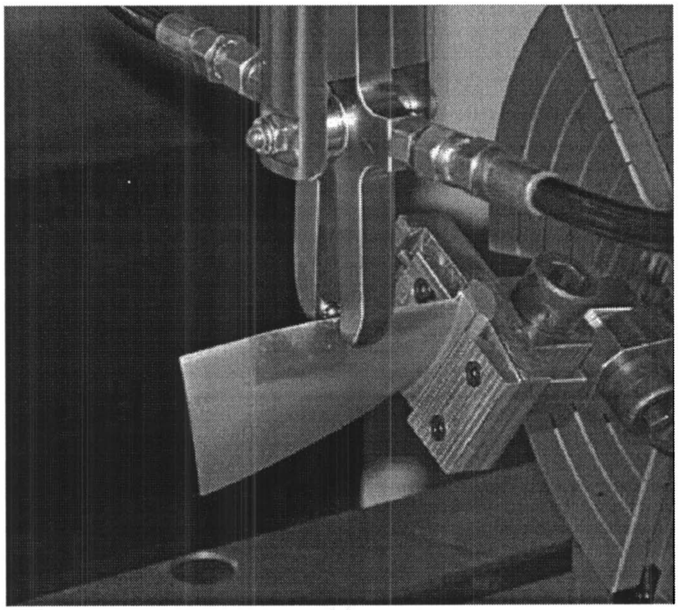
The value of the x-ray elastic constants required to calculate the macroscopic residual stress from the crystal lattice strain were determined in accordance with ASTM E1426-91 [29]. Systematic errors were monitored per ASTM specification E915.

The $K\alpha_1$ peak breadth was calculated from the Pearson VII function fit used for peak location during macroscopic stress measurement [30]. The percent cold work was calculated using an empirical relationship established between the material cold working (true plastic strain) and the $K\alpha_1$ line broadening [31]. The percent cold work is a scalar quantity, taken to be the true plastic strain necessary to produce the diffraction peak width measured based on the empirical relationship.

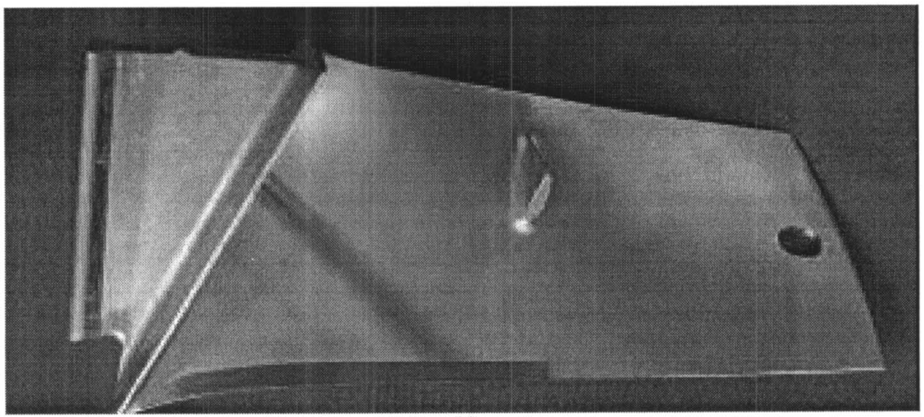
Thermal Stability of LPB Residual Stress Distributions Surface enhancement methods introduce a layer of compression on the surface of a component to enhance HCF performance or to mitigate the fatigue or stress corrosion debit caused by pitting corrosion, FOD, or other surface damage. Historically, shot peening (SP) has been widely used to significantly improve HCF performance. However, the heavy surface cold work, surface roughness, and potential surface damage associated with SP limit the range of application. The high degree of cold work produced by SP leads to both thermal and mechanical relaxation [5,7], and may leave the surfaces more prone to corrosion related damage. In contrast, LPB inherently produces little cold work in developing a deep layer high compression. LPB processing parameters (including ball size, type, applied pressure/vertical displacement, lateral velocity, feed, work piece mechanical properties, etc.,) are developed for a given material and geometry using x-ray diffraction measurement of residual stress and cold work distributions. Discussion of the LPB parameter development is beyond the scope of this paper.

Depending on the LPB process parameters and material characteristics, the residual stress state and the corresponding percent cold work (both depth profile and surface distribution) vary as functions of position. A large database has been developed to document these distributions for a variety of structural engineering materials of interest to aerospace, automotive, nuclear, and biomedical applications. A typical set of residual stress and corresponding percent cold work depth profiles for IN718 are shown in Fig. 3. The effects of thermal exposure to 525 and 600°C on the residual stress profiles are included. For the sake of comparison, typical results from shot peened (SP) surfaces given the same thermal exposure are included in these figures. Similar thermal relaxation studies have been conducted in many other material systems of interest to the turbine engine industry. Figure 4 shows the relaxation effects in LPB processed Ti-6Al-4V after exposure to 425°C.

In the IN718 example, it is seen that the LPB process achieved compression to a depth exceeding 1 mm (0.04 in.), reaching a maximum of nominally −1100 MPa (−159.6 ksi) at nominally 250 μm (0.01 in.) with less than 5% cold work through the compressive zone. In contrast, SP produced a depth of compression less than 250 μm (0.01 in.) with over 60% cold work on the surface. Retention of surface compression is important in HCF where suppression of surface fatigue initiation typically governs life. After 100 hours of exposure to 525°C, the LPB processed material showed very little relaxation of residual stress at the surface, while the SP processed material lost surface compression from over −800 MPa (−116 ksi) to under −100 MPa (14.5 ksi). This stress relaxation is attributed to the heavy cold work experienced by SP, leading to rapid recovery and stress relaxation during



(a)



(b)

Fig. 2 (a) 17-4 PH compressor blade being processed on the leading edge (LE) with an LPB caliper tool to improve FOD tolerance; (b) Ti-6-4 fan blade LPB processed along the lower LE. The hole near tip was made for HCF testing in cantilever bending (see Fig. 8).

thermal exposure [5]. Similar thermal relaxation data have been obtained for LPB and SP in a number of Ti, Fe and Al alloys and similar response to thermal exposure have been observed [6,7]. Using larger tools and higher pressures, compression up to 8 mm (0.3 in.) depth has recently been achieved in stainless steel welded plates.

Fatigue Testing. HCF tests were conducted on specially designed test specimens to simulate the service conditions of the components. For this purpose, thick section [9.5 mm (0.37 in.)] specimens with trapezoidal cross section (designed for testing highly compressive surfaces) or a blade-edge feature cross section were tested under a four-point bending load with the capability to vary applied R ratios ($\sigma_{\min}/\sigma_{\max}$). Crack propagation for the thick section samples was from the surface through the depth of the compressive layer. Blade-edge samples were processed to produce through-thickness compression and tested with the crack propagating from the edge, perpendicular to the processed surfaces. In most instances, an R ratio of 0.1 was used to simulate the high mean load common to rotating turbine components. HCF tests were run at a frequency of 30 Hz to either failure or a run-out condition of 2.5×10^6 cycles.

Engine operating conditions were simulated in several ways. For example, typical FOD conditions were simulated by either pressing an indentation, machining a notch, or using electro-discharge machining (EDM). To simulate exposure to engine temperatures, specimens were typically exposed to engine operating temperatures for a fixed time before HCF testing to allow thermal relaxation. Active corrosion was simulated by wrapping pads soaked in corroding media around the samples for the duration of the test. Fretting fatigue was simulated using a specially designed ‘‘bridge’’ fretting fixture with variable fretting normal loads (Fig. 13).

Effect of LPB on HCF Performance and FOD Tolerance. The thick section HCF results for IN718 presented in Fig. 5 show a substantial increase in the HCF endurance limit, or fatigue strength at 2×10^6 cycles for LPB over shot peening for either 525 or 600°C exposure for 100 h. The similar fatigue performance for shot peening followed by either 525 or 600°C exposure is attributed to the near uniform relaxation of the surface compression seen in Fig. 3 after exposure to either temperature. The endurance limit is typically associated with surface residual compression governing the initiation of fatigue cracks while fatigue strength in the finite life regime is dominated by crack growth through the depth of the compressive layer left by surface enhancement.

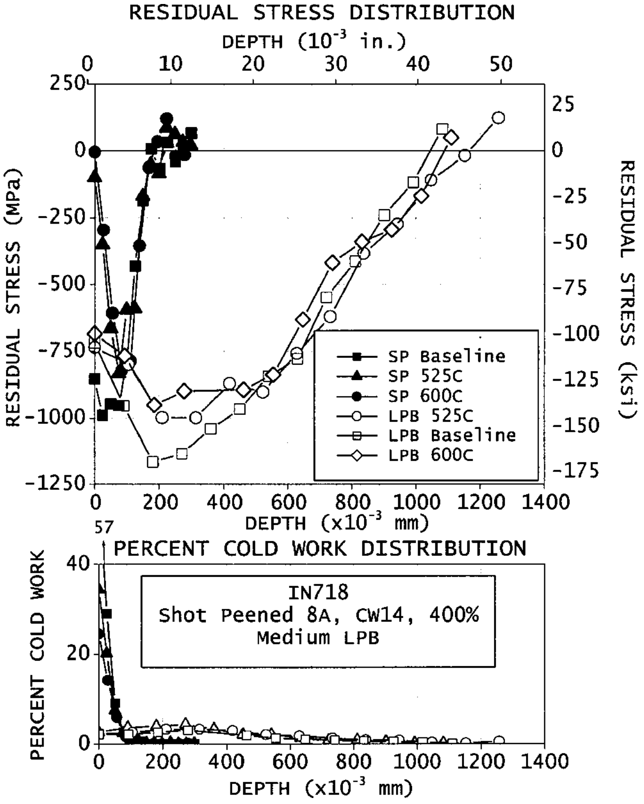


Fig. 3 Residual stress and cold work depth profiles for IN718 before and after exposure to service temperature of 525 and 600°C for 100 h

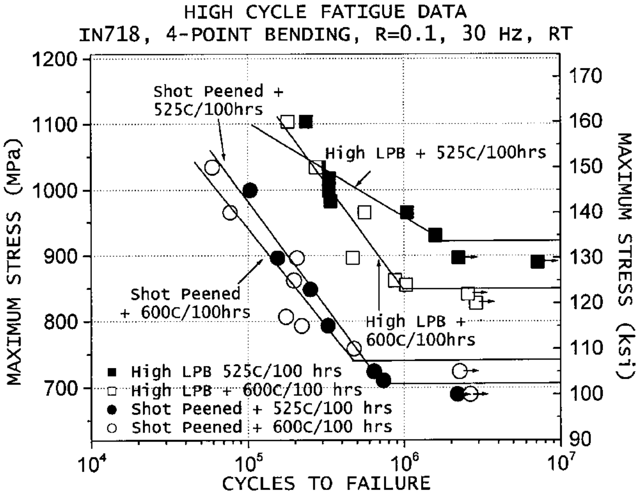


Fig. 5 High cycle fatigue performance of shot peened (8 A, 400%) and LPB processed thick section IN718 after 100 h, exposures at 525 and 600°C

The difference in the ability of the two surface enhancement methods to resist FOD either in the form of a single indentation or a sharp notch is shown in Fig. 6. The endurance limit is reduced from nominally 700 to 300 MPa (101. to 43.5 ksi) by either form of damage. The deep compressive layer produced by LPB is more effective in retarding crack growth, even after thermal exposure, because of the minimal thermal stress relaxation and the greater depth of the compressive layer. Although considerable scatter is evident in the LPB+FOD data (which is attributed to variability in the FOD damage) all of the specimens treated by LPB have fatigue strengths and lives superior to that of shot peening even without FOD.

HCF data for thick section Ti-6Al-4V shows similar trends in Fig. 7. The HCF results presented in this figure show a 38% increase in the endurance limit for LPB [>620 MPa (>90 ksi)] compared to shot peening [~448 MPa (~65 ksi)] after exposure to 425°C (795°F) for 10 h. The increased endurance limit after surface enhancement is generally associated with surface compression delaying the initiation of fatigue cracks at the surface. The reduced HCF strength of the highly cold worked shot peened condition is attributed to the complete loss of surface compression after even a brief elevated temperature exposure.

Improved Damage Tolerance in Aero Engine Components. LSP of the leading edges of fan and compressor blades improves damage tolerance, in terms of FOD depth, by at least an oder of magnitude. The high magnitude through-thickness compression achieved by LSP retards both fatigue crack initiation and growth, but requires expensive specialized laser systems. LPB has been

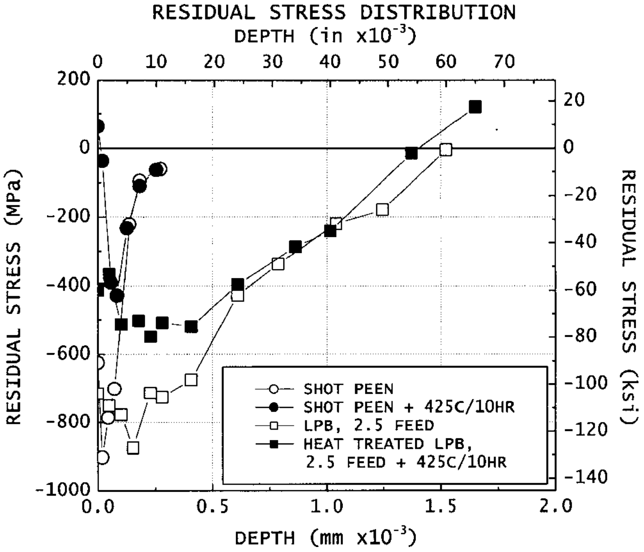


Fig. 4 Thermal relaxation of shot peened (8 A, 400%) and LPB processed Ti-6Al-4V after 795°F (425°C) for 10 h

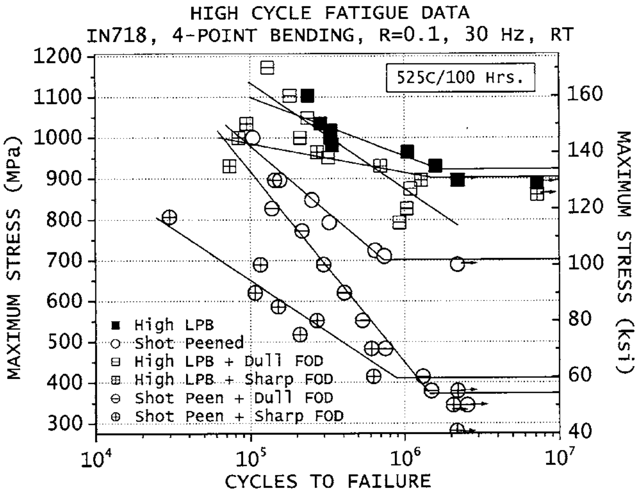


Fig. 6 HCF tolerance of dull and sharp FOD after elevated temperature exposure simulating engine environments for shot peened (8 A, 400%) and low plasticity burnished (LPB) thick section IN 718

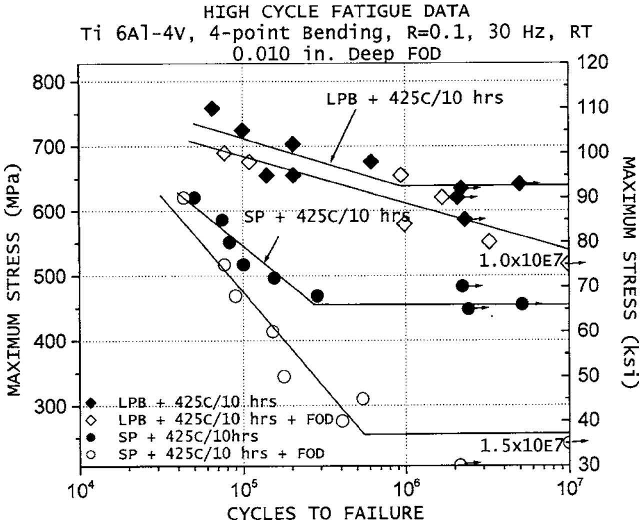


Fig. 7 HCF performance of thick section Ti-6-4 after SP and LPB surface treatments with and without the presence of a 250 μm (0.01 in.) FOD after exposure to service temperatures

investigated as a means of inducing through-thickness leading edge compression for improved damage tolerance in first stage Ti-6-4 fan blades. Fatigue performace was assessed by testing actual fan blades with controlled depth simulated leading edge FOD.

To establish the appropriate depth of simulated FOD, size and location distributions of service generated FOD were documented for used blades. Service induced FOD on the leading edge ranged from 0.005 mm (0.0002 in.) to less than 0.50 mm (0.02 in.) deep. Ninety percent of FOD depths observed were less than 0.05 mm (0.002 in.). Blades are currently removed for re-work or replacement if FOD 0.13 mm (0.005 in.) deep is found durng inspection. Conservatively, 0.50 mm was chosen as the minimum depth for FOD to be induced in blades for this study. Simulated FOD 0.5 mm (0.02 in.) and 1.2 mm (0.05 in.) in depth was created by machining 60° "V" notches in blades in the leading edge location at which maximum stress occurred during fatigue loading.

LPB processing CNC code was designed and LPB parameters developed to create through-thickness compression of nominally −689 MPa (−100 ksi) from the blade leading edge to a nominal distance of 6.4 mm (0.25 in.) chord-wise from the leading edge. The extent of the LPB zone running along the leading edge, from the midspan damper to the platform, may be seen clearly as the region of improved surface finish in Fig. 2(b).

Fatigue comparisons were made using used blades as test specimens cantilever loaded at $R=0.1$ to simulate the high mean load imposed on the leading edge during service in a rotating engine, as shown in Fig. 8. The S-N curves developed are shown in Fig. 9. In the absence of FOD, LPB processing increased fatigue strength by nominally 207 MPa (30 ksi) over nonprocessed blades, and approached the material yield strength. In all but one instance, LPB processed blades tested without FOD at such high stress levels failed outside the most highly stressed LPB-processed area, usually in the dovetail region.

Introduction of 0.5 mm (0.02 in.) deep simulated FOD in non-LPB blades reduced the fatigue strength 65% from 689 to 241 MPa (100 to 35 ksi). Introduction of the same 0.5 mm (0.02 in) FOD after LPB processing resulted in fatigue strength equal to that of non-LPB processed blades without FOD. FOD 1.27 mm (0.05 in.) deep after LPB reduced the fatigue strength to 620 MPa (90 ksi), only 10% less than the strength of unprocessed blades.

Similarly, Fig. 10 shows the beneficial effects of LPB in mitigating HCF damage and improving FOD tolerance in a Ti-6-4 low

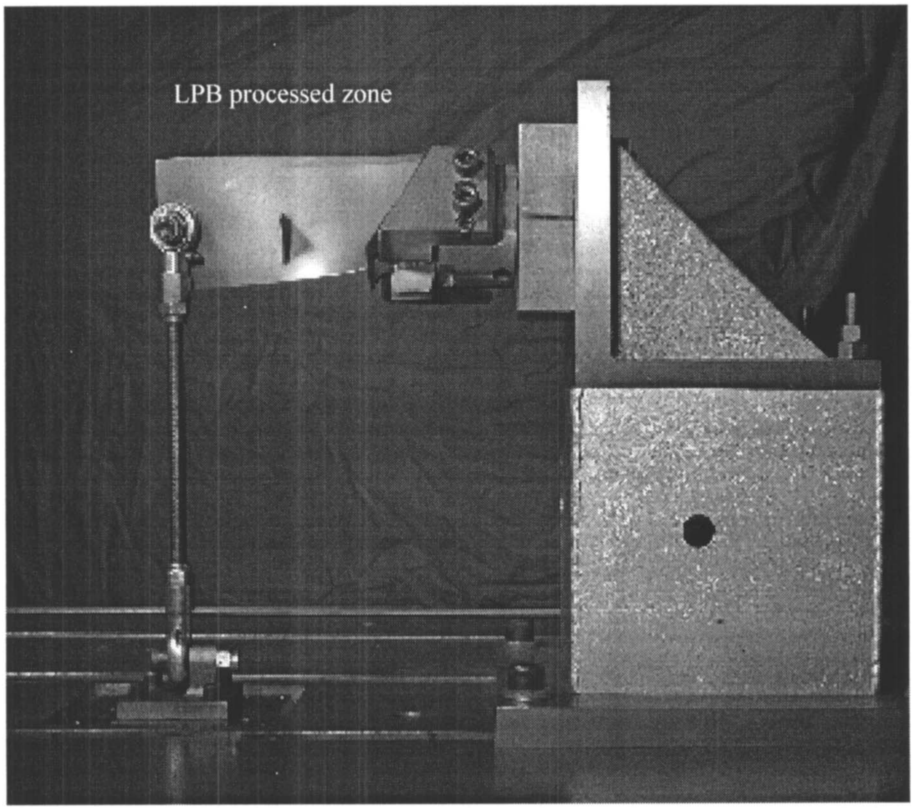


Fig. 8 Photograph showing the cantilever-loading fixture for fatigue testing of the first stage fan blade. Note the LPB patch along the (upper) leading edge of the blade.

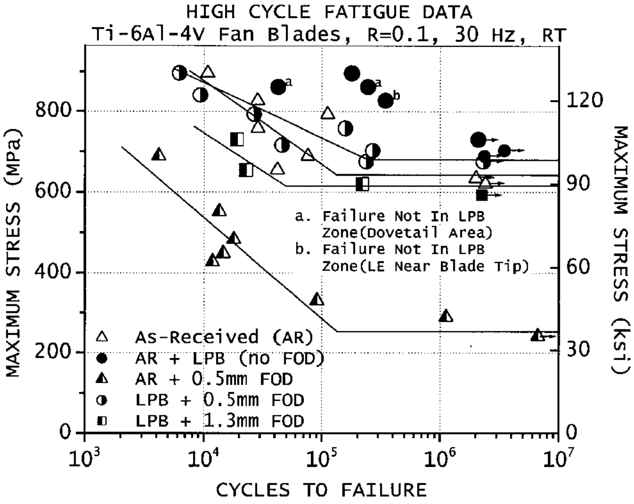


Fig. 9 Effect of LPB in mitigating HCF and FOD damage in Ti-6-4 fan blades

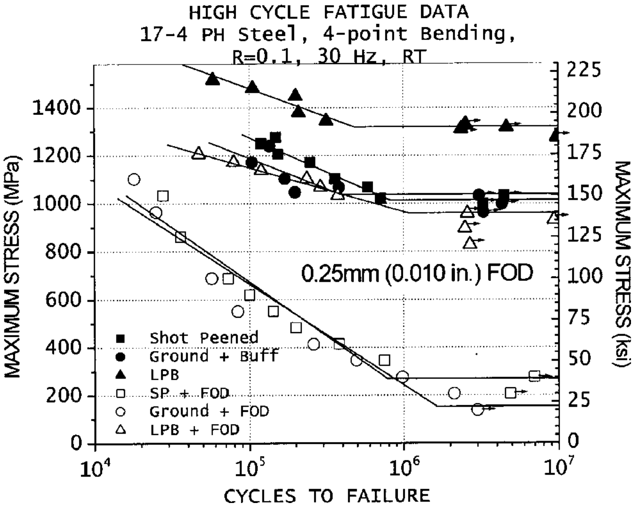


Fig. 11 HCF performance of thick-section 17-4 PH stainless steel with 0.25 mm (0.010 in.) FOD for base line, shot peened and LPB surfaces

pressure vane LPB processed to produce nominally −689 MPa (100 ksi) through-thickness compression on the loaded edge. The vanes were tested in cantilever loading conditions similar to the test methods used for the fan blades shown in Fig. 8. The fatigue strength of the LPB processed vanes [even in the presence of a 0.5 mm (0.02 in.) FOD] approached the tensile yield strength of the material.

The high compression developed by LPB improves fatigue strength with and without FOD by retarding both crack initiation and growth. However, the benefit is most evident with surface damage present. The through-thickness compression superimposed upon the applied tensile stress shifts the mean stress toward or into compression, resulting in a higher allowed alternating stress. Fracture mechanics based fatigue life analysis using the AFGROW code confirms that for an edge crack growing in a thin plate through a zone of through-thickness compression, the endurance limit is nominally equal to the magnitude of compression present, and the depth of damage (crack) tolerated is approximately equal to the width of the zone of through-thickness compression (measured from the edge.)

Damage Tolerance Improvements in 17-4 PH Stainless Steel. Figure 11 shows the HCF performance of thick section 17-4 PH stainless steel. In this plot, the base-line data (low-stress ground and buffed surface condition) is compared with HCF data for shot peened and LPB conditions. The base-line fatigue strength is 930 MPa (135 ksi), only slightly lower than the 965 MPa (140 ksi) strength of the shot peened condition. In contrast, LPB produced a fatigue strength (in bending) of 1240 MPa (180 ksi), exceeding the material's tensile yield strength [1033 MPa (150 ksi)] due to the introduction of residual compression on the order of the yield strength.

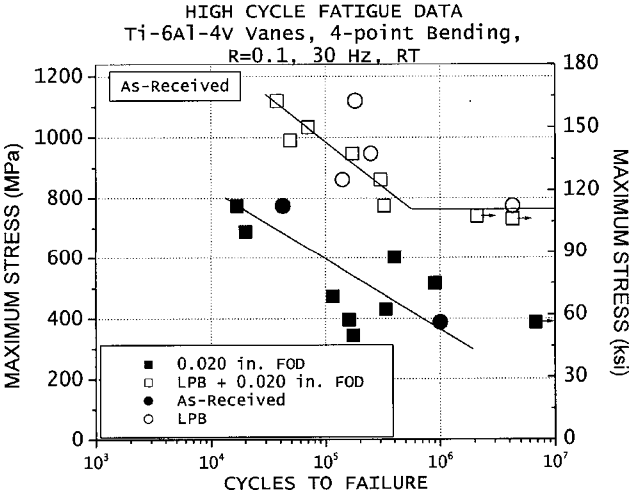


Fig. 10 Effect of LPB in mitigating HCF and FOD damage in a Ti-6-4 low pressure vane (LPV)

The HCF performance without LPB dropped dramatically when 0.25 mm (0.01 in.) deep by 0.76 mm (0.03 in.) long FOD was introduced by EDM. As seen in Fig. 11, SP and the base-line fatigue strengths with FOD were only 276 MPa (40 ksi) and 172 MPa (25 ksi), respectively. With the same FOD, the fatigue strength of LPB surface treated specimens was nominally 1033 MPa (150 ksi), comparable to the strength of the base-line material without FOD.

Mitigation of Corrosion Damage in Turbine Engine Alloys. Turbine engine components can be seriously affected by the presence of corrosion environments leading to pitting and significant debit in HCF lives. Salt spray corrosion pits are a common fatigue crack initiation site in ferritic stainless steels. Salt corrosion pitting can result from exposure to a marine atmosphere or steam turbine environments. Pitting depths, and the corresponding stress intensity factor, depend upon the time of exposure, temperature, and the service environment of the turbine engine. Salt pit corrosion typically reduces the endurance limit to nominally half of the uncorroded value. The effects of corrosion are often further exacerbated by the presence of FOD or erosion.

The effect of active corrosion (during fatigue cycling) on HCF behavior was monitored by exposing the thick section bend specimens to a corrosive medium during fatigue tests. A chemical-free absorbent pad soaked in an acidic salt solution (3.5 wt % NaCl, *p*H 3.5) was taped to the gauge section of the specimens and sealed to prevent evaporation with a thin plastic film. All other test conditions were as described earlier for studying thick-section HCF behavior. The effects of LPB on the active corrosion fatigue performance of Custom 450 stainless steel are presented in Fig. 12.

The low-stress ground base-line fatigue strength, tested in acid salt solution without FOD, was nominally 689 MPa (100 ksi). The base-line fatigue strength with 0.25 mm (0.01 in.) FOD was only 69 MPa (10 ksi). In comparison, the LPB treated samples had fatigue strengths of over 1102 MPa (160 ksi) in the corrosive

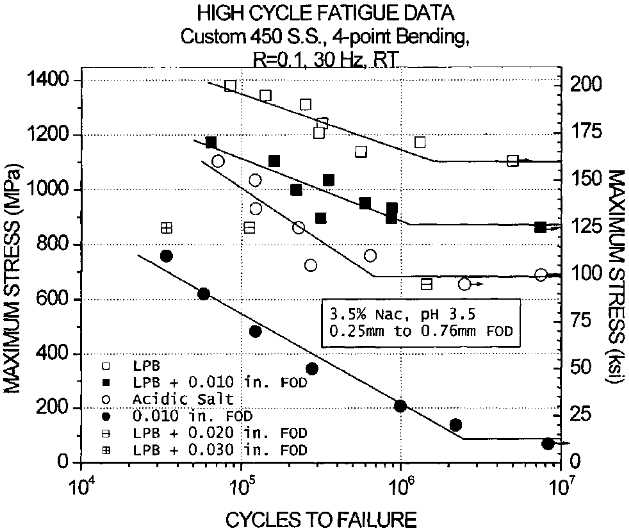


Fig. 12 Corrosion fatigue performance of Custom 450 stainless steel with FOD in acidic salt solution and FOD from 0.25 mm (0.010 in.) to 0.76 mm (0.030 in.)

environment alone, and 861 MPa (125 ksi), well above the baseline condition without FOD, with 0.25 mm (0.01 in.) FOD. Although full S-N curves were not generated, when FOD depth was increased to 0.50 mm (0.02 in.) and 0.76 mm (0.03 in.), the further debit in HCF performance was minimal.

Mitigation of Fretting Fatigue Damage in Ti-6Al-4V. Fretting damage and the resulting HCF performance debit can reduce the HCF strength by 50% and limit the life of turbine engine fan and compressor dovetail joints. Fretting occurs where component surfaces are pressed into physical contact under loading conditions that produce small relative movement. Shallow shear stress cracks initiated at the edges of the fretting scars can grow in mode I to failure.

Controlled fretting damage HCF tests were conducted on thick section Ti-6-4 to determine whether a deep layer of high residual compression produced by LPB could arrest the fretting induced shear cracking and prevent mode I propagation to failure, improving the fretting fatigue strength. Fretting damage was introduced into two sets of HCF samples both by pre-fretting the sample prior to fatigue testing (to introduce prior damage), and by fretting during testing. Fretting damage was produced by pressing two Ti-6-4 alloy cylindrical rods with a 6.3 mm (0.25 in.) diameter into the active gage surface of the fatigue specimen gage section using a bridge device to hold the rods. The rods were pressed into the fatigue specimen with a constant load of approximately 33.7 N (150 lbs) (i.e., 16.8 N per rod), producing a nominal contact stress of approximately 462 MPa (67 ksi). The fretting fixture was instrumented with strain gages and calibrated to record the normal load during testing. Figure 13 shows a photo of the fretting fixture positioned on a fatigue specimen.

The fretting fatigue results for thick section Ti-6-4 are shown in Fig. 14. Specimens were surface treated, exposed to simulated engine temperatures (375°C for 10 h) and then subjected to fretting fatigue at room temperature. The fatigue strength for electropolished base-line Ti-6-4 tested at $R=0.1$ in four-point bending is nominally 482 MPa (70 ksi), which decreased drastically to about 172 MPa (25 ksi) when subjected to fretting. Compressive residual stresses from shot peening provide significant benefit in restoring the fatigue strength to 413 MPa (60 ksi), nearly the base-line strength. In comparison, LPB produced nearly full restoration at 482 MPa (70 ksi). All but two LPB specimens tested at high stress levels [758 and 827 MPa (110 and 120 ksi)] failed outside of the LPB zone or sub-surface, with crack initiation from defects just below the compression zone, and not in the fretting scars.

Summary

The life of turbine engine and power systems is limited by fatigue, FOD, corrosion fatigue and fretting fatigue. Even small FOD, corrosion pits, or other surface damage can nucleate fatigue failures with catastrophic consequences. The cost of inspection and maintenance required to avoid fatigue failures is estimated in

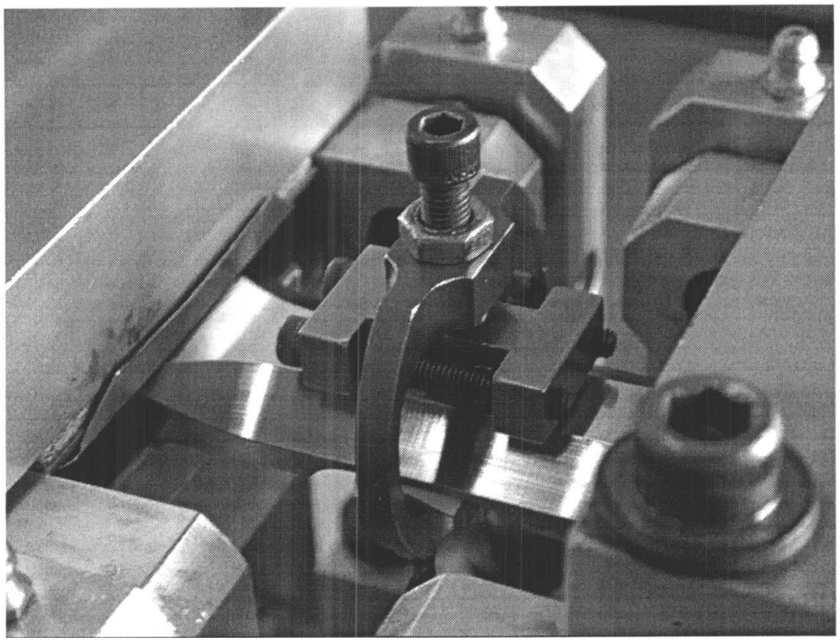


Fig. 13 Photograph of the fretting fixture showing the top two fretting rods clamped onto the top surface of the thick section HCF specimen

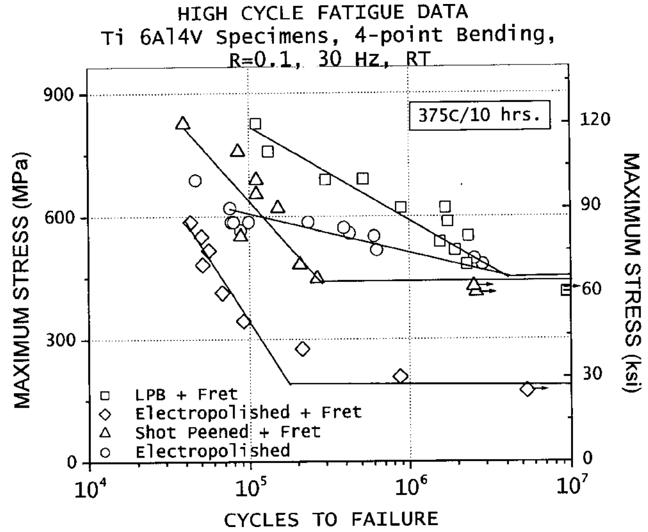


Fig. 14 Ti-6Al-4V fretting fatigue data for shot peened and LPB surface treatments

the billions of dollars annually. Design or material changes generally provide only limited fatigue improvement at huge cost. Surface enhancement by LPB provides a thermally stable deep layer of compressive residual stress that can dramatically improve fatigue performance of turbine alloys without altering either alloy or design.

As summarized in this paper, the deep compression from LPB has been shown to resist thermal relaxation at turbine operating temperatures far better than conventional shot peening. The deep layer of compression has been shown to completely mitigate FOD up to 0.25 mm (0.01 in.) deep in Ti-6-4, IN718 and 17-4 PH. Damage 1.27 mm (0.05 in.) deep can be tolerated in Ti-6-4 blade edges. The corrosion fatigue strength of Carpenter 450 and fretting fatigue performance of Ti-6-4 can be substantially improved. Performance studies of other alloys are in process with comparable initial results.

LPB can be performed on conventional CNC machine tools in a conventional machine shop environment either in original manufacturing or during overhaul and repair. Residual stress distributions designed to cancel applied tension can be induced into the component surface to provide optimal fatigue performance at costs comparable to conventional machining. LPB offers the turbine designer a means, in addition to material selection and design, to improve turbine component performance and life at reduced operating costs.

Acknowledgment

Support for the development of the data presented here was provided under SBIR programs through NASA, Glenn Research Center, under the direction of Dr. Timothy Gabb; Wright Patterson AFB Materials Lab under the direction of Michael Shepard, the US Navy, NAVAIR, under the direction of Ravi Ravindranath, and with internal research support from Lambda Research. The financial support and technical guidance provided is gratefully acknowledged and much appreciated.

Nomenclature

LPB = Low plasticity burnishing
SP = Shot peening
HCF = High cycle fatigue
EDM = Electrical-discharge machining
FOD = Foreign object damage
CW = Cold work
RS = Residual stress
XRD = X-ray diffraction

References

[1] U.S. Patents 5,826,453 (October 1998), 6,415,486 B1 (Jul. 2002) other US and foreign patents pending.
[2] 2002, “Low Plasticity Burnishing,” *NASA Tech Briefs*, Aug., 50 p.
[3] Hogan, B., ed., 2001, “Longer Life With Low-Plasticity Burnishing,” *Manufacturing Engineering*, SME, pp. 34–38.
[4] Gabb, T., Telesman, J., Kantzos, P., and Prevéy, P., 2002, “Surface Enhancement of Metallic Materials,” *Advanced Materials & Processes*, ASM, ed., Peg Hunt, Jan., pp. 69–72.
[5] Prevéy, P., 2000, “The Effect of Cold Work on the Thermal Stability of Residual Compression in Surface Enhanced IN718,” Proc. 20th ASM Materials Solutions Conf., St. Louis, MO, Oct. 10–12.
[6] Prevéy, P. et al., 2000, “FOD Resistance and Fatigue Crack Arrest in Low Plasticity Burnished IN718,” Proc. 5th National HCF Conf.
[7] Prevéy, P. et al., 2001, “The Effect of Low Plasticity Burnishing on the HCF Performance and FOD Resistance of Ti-6Al-4V,” Proc. 6th Natl. Turbine Engine HCF Conf., Jacksonville, FL, March 5–8.
[8] Belozerov, V. V., Makhatilova, A. I., Turovskii, M. L., and Shifrin, I. M., 1986, “Increasing the Fatigue Resistance of High-Strength Steel Using Bulk and Surface Treatments,” Met. Sci. Heat Treat., **28**(7-8), pp. 565–569.
[9] Stepurenko, V. T. et al., 1976, Prot. Met., **12**, pp. 386–389.
[10] Papshev, D. D., and Golubev, Yu G., 1972, “Effectiveness of Surface Work-Hardening of Titanium Alloy Components,” Russ. Eng. J., **52**(4), pp. 48–51.
[11] Freid, M. Kh. et al., 1994, Prot. Met., **20**, pp. 263–265.
[12] Belkin, L. M. et al., 1984, “Surface Strengthening of Plane Parts by Plastic Deformation,” Sov. Eng. Res., **4**(9), pp. 30–32.
[13] Belkin, L. M., 1983, “Survivability of Flat Parts Strengthened by Surface Plastic Deformation,” Sov. Mater. Sci., **19**(3), pp. 225–228.
[14] Fattouh, M. et al., 1988, Wear, **127**, pp. 123–137.
[15] Loh, N. H. et al., 1989, Wear, **129**, pp. 235–243.
[16] Loh, N. H. et al., 1993, Precis. Eng., **15**, pp. 100–105.
[17] Kotiveerachari, B., and Murty, R. L., 1985, Int. J. Prod. Res., **23**, pp. 499–521.
[18] Hills, D. A. et al., 1979, Proc. Int. Conference Wear of Materials, ASME, New York, NY, pp. 396–402.
[19] Braham, S., and Frelat, J., 1993, *Proc. Computer Methods and Exp. Meas. for Surface Treatment Effects*, Computational Mechanics, Southampton, UK, pp. 255–264.
[20] Zinn, W., and Scholtes, B., 1999, “Mechanical Surface Treatments of Lightweight Materials—Effects on Fatigue Strength and Near-Surface Microstructures,” J. Mater. Eng. Perform., **8**, pp. 145–151.
[21] Altenberger, I. et al., 1999, “Cyclic Deformation and Near Surface Microstructures of Shot Peened or Deep Rolled Austenitic Stainless Steel AISI 304,” Mater. Sci. Eng., A, **264**, pp. 1–16.
[22] Drechsler, A. et al., 1998, “Mechanical Surface Treatments of Ti-10V-2Fe-3Al for Improved Fatigue Resistance,” Mater. Sci. Eng., A, **243**, pp. 217–220.
[23] Hilley, M. E., ed., 1971, *Residual Stress Measurement by X-Ray Diffraction, SAE J784a*, Society of Auto. Eng., Warrendale, PA.
[24] Noyan, I. C., and Cohen, J. B., 1987, *Residual Stress Measurement by Diffraction and Interpretation*, Springer, New York.
[25] Cullity, B. D., 1978, *Elements of X-ray Diffraction*, 2nd ed., Addison-Wesley Reading, MA, pp. 447–476.
[26] Prevéy, P. S., 1986, “X-Ray Diffraction Residual Stress Techniques,” *Metals Handbook*, ASME, Metals Park, OH, Vol. 10, pp. 380–392.
[27] Koistinen, D. P., and Marburger, R. E., 1964, Trans. ASM, **67**, p. 67.
[28] Moore, M. G., and Evans, W. P., 1958, “Mathematical Correction for Stress in Removed Layers in X-Ray Diffraction Residual Stress Analysis,” SAE Trans., **66**, pp. 340–345.
[29] Prevéy, P. S., 1977, “A Method of Determining Elastic Properties of Alloys in Selected Crystallographic Directions for X-Ray Diffraction Residual Stress Measurement,” *Advances in X-Ray Analysis*, Plenum, New York, 1977, Vol. 20, pp. 345–354.
[30] Prevéy, P. S., 1986, “The Use of Pearson VII Functions in X-Ray Diffraction Residual Stress Measurement,” *Advances in X-Ray Analysis*, Plenum, New York, Vol. 29, pp. 103–112.
[31] Prevéy, P. S., 1987, “The Measurement of Residual Stress and Cold Work Distributions in Nickel Base Alloys,” *Residual Stress in Design, Process and Material Selection*, ASM, Metals Park, OH.

Anders Angantyr
ALSTOM Power Sweden AB,
72176 Västerås,
Sweden
e-mail: anders.angantyr@power.alstom.com

Jan-Olov Aidanpää
Division of Computer Aided Design,
Department of Applied Physics and Mechanical Engineering,
Luleå University of Technology,
97187 Luleå,
Sweden
e-mail: joa@cad.luth.se

Constrained Optimization of Gas Turbine Tilting Pad Bearing Designs

This paper presents the constrained optimization of the tilting pad bearing design on a gas turbine rotor system. A real coded genetic algorithm with a robust constraint handling technique is used as the optimization method. The objective is to develop a formulation of the optimization problem for the late bearing design of a complex rotor-bearing system. Furthermore, the usefulness of the search method is evaluated on a difficult problem. The effects considered are power loss and limiting temperatures in the bearings as well as the dynamics at the system level, i.e., stability and unbalance responses. The design variables are the bearing widths and radial clearances. A nominal design is the basis for comparison of the optimal solution found. An initial numerical experiment shows that finding a solution that fulfills all the constraints for the system design is likely impossible. Still, the optimization shows the possibility of finding a solution resulting in a reduced power loss while not violating any of the constraints more than the nominal design. Furthermore, the result also shows that the used search method and constraint handling technique works on this difficult problem. [DOI: 10.1115/1.2179463]

1 Introduction

During a late design stage, much of the dynamics of a rotor-bearing system are already determined by earlier design decisions. At this point it may, for example, be difficult to change the rotor geometry since this affects the function of the machine. However, the bearing design is easily changed even at a late stage. The bearing design has an important impact on the system dynamics, e.g., stability. Its design is often an iterative process and opens the possibilities for efficient use of optimization methods.

Optimizations of a single bearing design have been performed in many papers, see, for example, [1]. Optimizations of the dynamics of rotor-bearing systems with rotor dimension parameters and bearing stiffness coefficients as design variables have also been studied, e.g., [2,3]. The authors of this paper believe that optimizing bearing design and system dynamics separately will not result in optimal system performance. This is because the dynamics of the whole system is closely coupled to the bearing design. Recent studies with bearing design parameters as design variables and system dynamics as the target have been done in [4–6]. In [4,5], the optimization was performed on a single rotor supported on two bearings (two-lobe type), with no information being provided on the bearing model. The rotor weight, natural frequency, and damping of a single mode were considered. In [6], a generator with three bearings was studied. A bearing database was used, and several modes were considered.

This paper presents the optimization of four titling pad bearings in a rotor system, including gas turbine, intermediate shaft, and gear. The objective is to minimize power loss with constraints on bearing temperatures and stability. The amplitudes for several unbalance response cases are also introduced as constraints. The bearing widths and radial clearances are chosen as design variables since these may be subject to changes in the late stage of the design. Nonlinear bearing analyses are performed within the optimization loop. This is a highly constrained problem where nothing is known about the shape of the objective function or the constraints. The global optimum is of interest, and local search efficiency is of secondary importance. Therefore, a real coded genetic algorithm (GA) [7] with the constraint handling method proposed by Angantyr et al. [8] is used. In [4–6,9,10], GAs were used in the design of simpler rotor-bearing systems. The main objective of this paper is to develop an optimization problem formulation for a case with an industrial degree of complexity. The second objective is to evaluate usefulness of the search method.

A well-developed and mature nominal design of the rotor-bearing system exists. However, the nominal design is not feasible since some of the constraints are unfulfilled. By definition, a feasible design is a design that satisfies all constraints. If a feasible design cannot be found, the paper should answer whether a better design than the nominal at least exists. The problem is, generally speaking, to be regarded as difficult. The problem formulation is specific for this particular case. However, it might inspire others working on similar problems. The used search method and constraint handling technique are generally and widely applicable and should therefore be of interest to a broad audience.

Bearing and system dynamics analyses are presented in Sec. 2. The objectives and constraints for the system design are defined in Sec. 3. An initial numerical experiment is explained in Sec. 4. The purpose of this experiment is to gain a better insight into the parameters possible effect on the system. The actual optimization problem is formulated in Sec. 5, and the results of the numerical experiment and optimization are given in Sec. 6. Finally, some concluding remarks are given.

2 Model and Analysis Approach

This section describes the dynamical analyses and models. A schematic sketch of the 43 MW gas turbine rotor-bearing system is found in Fig. 1. The compressor shaft is coupled via an intermediate shaft to the pinion wheel in the gear (left side in Fig. 1). The total weight of the rotor is 8800 kg, and the total polar inertia is 778 kg m^2. The gas turbine normally operates at 6608 rpm.

The bearing stiffness and damping properties are nonlinear, but a linear bearing model is used in the dynamical analyses of the rotor-bearing system. Hence, the bearings are described by eight stiffness and damping coefficients [11]. The rotor is discretized into 366 elements. For the dynamical analyses, an in-house code based on the transfer matrix method [12] and a sub-structuring method is used. The bearing supports 1 and 2 are described by

---

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received May 11, 2004; final manuscript received October 25, 2005. Review conducted by N. Arakere.

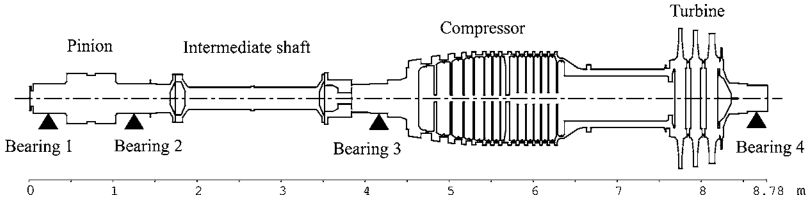


Fig. 1 Schematic sketch of rotor

simple mass-spring models. The gas turbine stator is described by its modal properties resulting from a finite element method (FEM) modal analysis of the complete stator. Hence, there is a coupling between bearing supports 3 and 4 in the dynamical analysis model. A modal damping ratio of 3% is assumed for the stator modes. Complex eigenvalue analyses (stability) and steady-state unbalance response analyses are performed in the optimization.

The bearings are hydrodynamical journal bearings of the tilting pad type. The bearings are analyzed under stationary conditions. For this purpose, a code based on ALP3T [13] is used. Typical output from the bearing analyses are eccentricity and attitude angle that define the center of the shaft. Other results are the linear stiffness and damping coefficients, power loss, maximum pad surface temperature, etc. In these analyses, a temperature-dependent viscosity model is used with similar fluid properties to an ISO VG32 oil. Figure 2 defines the bearing geometry.

The rotational direction is indicated by Ω. R defines the radius of each lobe; r defines the radius of the shaft. The preset distance for each lobe is t. Hence, the radial clearance is defined as $\Delta R = R - r - t$. The width of the bearing is B. The geometry of the lobes is defined by the angles φ_1, φ_2, and φ_3. The applied static load is F, and θ defines the direction of the force. For gravitational load only, the angle θ is zero. Observe that this is not the case if the gas turbine transfers a torque load. The contact forces in the gear then give rise to static loads in other directions. The geometry for the nominal design of the four bearings is given in Table 1.

The dynamical analyses of the system are linear, but the bearing analyses are nonlinear. The computational time required at the present date on a standard PC for the linear dynamical analyses is a few seconds, whereas the time required for a single bearing analysis is tens of seconds.

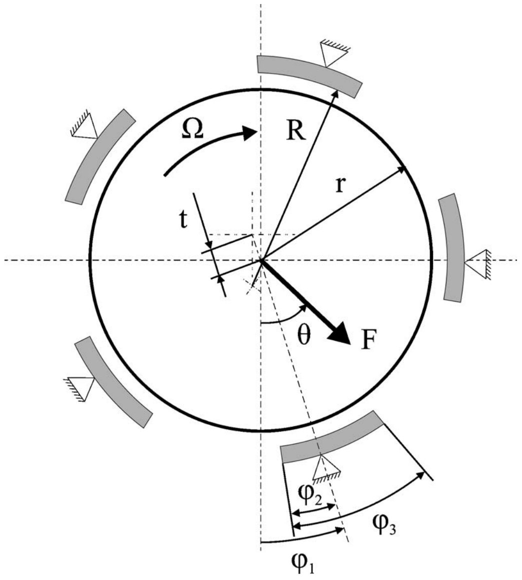


Fig. 2 Tilting pad bearing geometry

3 Design Objectives and Constraints

The objectives and constraints for the design of the rotor-bearing system are defined in this section. Clearly there exist several objectives for the system design that will not be considered here. The focus for this work is the optimization of the bearing designs. Therefore, only the objectives and constraints affected by the bearing designs will be defined.

3.1 Bearing Design. A target for the overall design of the gas turbine is to achieve a high efficiency. The bearings should obviously be designed for minimal power losses. This will be the objective of the problem formulation in Sec. 5.

It is important that the white metal on the pad surfaces in the bearings does not become too warm. Therefore, the bearing designs are constrained by the maximum allowed pad surface temperature. A function describing the feasibility with respect to the bearing surface temperature constraints is formulated as

$$T(\mathbf{x}) = \sum_{p=1}^{4} \max\left(0, \frac{T_p - T_p^L}{T_p^L}\right) \tag{1}$$

T_p is the maximum bearing surface temperature for bearing No. p at normal operating conditions. T_p^L is the maximum allowed bearing surface temperature for bearing No. p. Equation (1) gives a zero value for a design $\mathbf{x}$ having all temperatures below the specified limits. The vector of design variables $\mathbf{x}$ and maximum allowed bearing temperature T_p^L are defined later.

3.2 Stability. A necessary condition is for the system to be stable. Tilting pad bearings are known to have good stabilizing properties due to the small cross-coupling coefficients in the bearing stiffness matrix. Hence, stability should probably not be difficult to achieve. Still, this important criterion must be checked since an unstable system will certainly lead to failure and possibly severely damage the rotor-bearing system. Stability is determined by the complex eigenvalues of the system. These appear as conjugate pairs

$$\lambda_j = \alpha_j \pm i\omega_j \tag{2}$$

ω_j is the damped natural frequency for the jth eigenvalue. α_j is the growth factor and $\alpha_j > 0$ implies an unstable mode. Stability may also be expressed [4] by the damping ratio for the jth mode as

Table 1 Bearing geometry for nominal design

Bearing	r (mm)	B (mm)	ΔR (μm)	t (μm)	φ_1 (deg)	φ_2 (deg)	φ_3 (deg)
1 and 2	125	250	233	150	20	27	60
3	125	140	291	22	−36	24	60
4	110	154	255	23	−36	24	60

Table 2 Definition of the different stability cases

Case (k)	Operating conditions
1	40% speed and no torque load
2	60% speed and no torque load
3	80% speed and no torque load
4	100% speed and no torque load
5	100% speed and full torque load (normal operating conditions)

$$\zeta_j = -\frac{\alpha_j}{\sqrt{\alpha_j^2 + \omega_j^2}} \tag{3}$$

For a stable operation, the theoretical criterion is now $\zeta_j > 0 \forall j$. This must be ensured for all operating conditions. Therefore, the complex eigenvalues of the system must be stable for running up and normal operation. Since it is practically impossible to check stability conditions for the whole range up to normal speed, discrete cases k are defined and summarized in Table 2.

For 100% torque load, $F = 154$ kN and $\theta = -160$ deg for bearing 1 and $F = 148$ kN and $\theta = -157$ deg for bearing 2. The bearings at the gear are therefore subject to high load during normal operation.

For the real rotor, there may exist destabilizing effects, such as internal damping and fluid-induced forces, though these effects are not considered in the analysis. Therefore, a minimum required damping ratio is preferable. This is set to ζ^L. A function describing the feasibility regarding the operating case k is formulated as

$$C_k(\mathbf{x}) = \sum_{j=1}^{N} \max\left(0, \frac{\zeta^L - \zeta_{jk}}{\zeta^L}\right) \tag{4}$$

The number of complex eigenvalues in the frequency range of interest is N. The modal damping for the mode j and case k is ζ_{jk}. With this formulation a feasible design $\mathbf{x}$ regarding the case k satisfies $C_k(\mathbf{x}) = 0$. For an infeasible design, $C_k(\mathbf{x}) > 0$ holds.

3.3 Unbalance Response. For practical design purposes, the unbalance response analysis of a distributed real rotor system can be problematic due to the actual unbalance mass distribution along the rotor being generally unknown. The system should still be designed to have the unbalance response below standard values. In [14], an interesting approach based on singular value decomposition is proposed. This overcomes the problem with the unknown unbalance mass distribution and defines a general and conservative criterion that could be used for design purposes.

Another approach is to study the modal unbalance sensitivity [15]. When doing so, the interaction between modes is neglected. In this paper, a third approach is used. A set of different point mass unbalance cases is defined, as schematically shown in Fig. 3.

The rotor mass between two consecutive bearings is the unbalance mass m_ρ. This is 790 kg in case 1, 1050 kg in cases 2 and 3, and 6820 kg in cases 4 and 5. The radius to the unbalance mass is ρ and defined to fulfill the balancing grade G2.5 according to ISO 1904/1. In cases 3 and 5, the unbalance masses are shifted 180 deg in phase. The unbalance response is calculated for excitation frequencies from 85% to 120% of the operating speed. The bearings are analyzed for seven different speeds in this range and linear interpolation of the bearing coefficients is done for intermediate excitation frequencies.

The maximum amplitudes are checked at the bearing positions (1–4) and unbalance mass positions (5–10). Clearly, this approach would generate 50 constraints, i.e., $5 \times (4+6)$. To get a more manageable problem, constraint functions based on a weighted sum approach are formulated for each unbalance case.

Now let A_{mn} be the vibration amplitude (major axis of elliptical whirling orbit) for the unbalance case m and position defined by n.

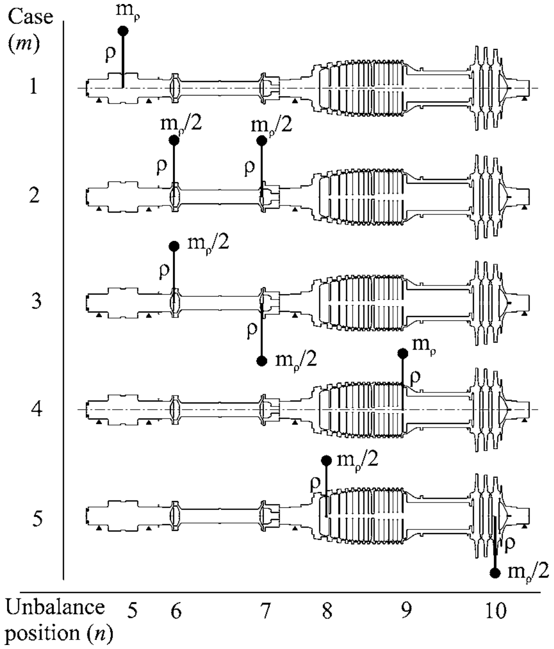


Fig. 3 Unbalance cases

The bearing positions correspond to $n = 1-4$. The unbalance positions indicated in Fig. 3 correspond to $n = 5-10$. For $n = 1-4$, A_{mn} is the maximum relative amplitude between the shaft and the bearing journals. Also let the maximum allowed vibration amplitude for unbalance m case and position n be A_{mn}^L. The function that describes the degree of feasibility with respect to the unbalance case m is then formulated as

$$U_m(\mathbf{x}) = \sum_{n=1}^{10} \max\left(0, \frac{A_{mn} - A_{mn}^L}{A_{mn}^L}\right) \tag{5}$$

If $U_m(\mathbf{x}) > 0$, the unbalance case m is violated for the design $\mathbf{x}$. A feasible design $\mathbf{x}$ satisfies $U_m(\mathbf{x}) = 0$.

3.4 Constraint Limits. The limits for the constraints are defined in this section. The maximum allowed temperature in bearings 1 and 2 is 110 °C and 100 °C for bearings 3 and 4, i.e., $\forall p \in \{1,2\}: T_p^L = 110$ °C and $\forall p \in \{3,4\}: T_p^L = 100$ °C in Eq. (1). The minimum allowed damping ratio for the complex eigenmodes is $\zeta^L = 2\%$. For the unbalance response, the relative amplitude in the bearings should not exceed 22 μm, i.e., $\forall m \in \{1, \ldots, 5\}$ and $n \in \{1, \ldots, 4\}: A_{mn}^L = 22$ μm. The amplitude at the unbalance positions should not exceed 33 μm, i.e., $\forall m \in \{1, \ldots 5\}$ and $n \in \{5, \ldots, 10\}: A_{mn}^L = 33$ μm.

The temperature and unbalance constraint are the most difficult to satisfy. The nominal design does not fulfill the constraints (Eq. (1) and Eq. (5)) with the above-given limits. However, if the constraint limits are set to the values as in Table 3, the nominal design becomes feasible.

Table 3 Limits required for feasibility of the nominal design

T_1^L (°C)	T_2^L (°C)	A_{26}^L (μm)	A_{27}^L (μm)	A_{43}^L (μm)	A_{46}^L (μm)	A_{47}^L (μm)
115	111	41.4	58.6	35.3	43.2	41.2

4 Numerical Experiment

To gain some insight into the parameters effect on the studied rotor-bearing system, an initial numerical experiment is performed. Although there are a wide variety of methods on how to design experiments, Montgomery [16], a simple but informative method is chosen. Solutions are generated randomly with uniform distribution within the limits specified in Eq. (6). The chosen factors for variation are the bearing widths and radial clearances. The ranges for the variation in the factors are

$$
\begin{gathered}
100 \text{ mm} \leqslant \mathrm{B}_1, \quad \mathrm{B}_2,\ \mathrm{B}_3 \leqslant 250 \text{ mm} \\
88 \text{ mm} \leqslant \mathrm{B}_4 \leqslant 220 \text{ mm} \\
138\ \mu\text{m} \leqslant \Delta\mathrm{R}_1, \quad \Delta\mathrm{R}_2 \leqslant 250\ \mu\text{m} \\
138\ \mu\text{m} \leqslant \Delta\mathrm{R}_3 \leqslant 310\ \mu\text{m} \\
121\ \mu\text{m} \leqslant \Delta R_4 \leqslant 270\ \mu\text{m} \qquad (6)
\end{gathered}
$$

The indices correspond to the bearing position. The choice of factors for the numerical experiment is based on the fact that these are the only possible factors to change at a late design stage. Furthermore, it is known from experience that these factors (at various amounts) affect the objectives and constraints discussed in Sec. 3. The result of this experiment is discussed in Sec. 6.

5 Optimization Problem

The subject of optimization is the design of the bearings. Since these are parts of the system, each bearing cannot be optimized separately. The objectives and constraints for the whole system design must be considered. One possibility may be to divide the overall problem into a top-down hierarchy of subproblems. The Target Cascading method [17] is an example of how to formulate and solve problems with this approach. In this case, the subproblems (bearing designs) are tightly coupled via numerous implicit constraints. Although the objectives for the subsystem level are quite obvious, they would be difficult to achieve. Therefore, a classical formulation of the optimization problem is used here. The optimization problem is stated as

$$
\begin{gathered}
\text{Minimize } P(\mathbf{x}) \\
\text{subject to} \\
T(\mathbf{x}) = 0 \\
C_k(\mathbf{x}) = 0 \quad \text{for } k = 1-5, \\
U_m(\mathbf{x}) = 0 \quad \text{for } m = 1-5. \qquad (7)
\end{gathered}
$$

$P(\mathbf{x})$ is the total power loss in the bearings under operating conditions. The functions in the constraints are defined by Eqs. (1), (4), and (5). The design variables are the same as the factors in the initial numerical experiment. Hence, the vector of design variables is $\mathbf{x}=[B_1,\Delta R_1,B_2,\Delta R_2,B_3,\Delta R_3,B_4,\Delta R_4]$. The side constraints of the design variables (upper and lower limits) are given in Eq. (6).

Nothing is known about the shapes of the objective and constraint functions. Since the global optimum is of interest, a robust global search method is a preferable optimization algorithm. Therefore, a real-coded GA is chosen. A good introduction to real-coded GAs is the book by Gen and Cheng [7]. A frequently used method to handle constraints is by penalty functions [18]. A drawback of these methods is often that problem dependent penalty coefficients have to be specified. In [8], Angantyr et al. proposed a robust and generic constraint-handling method requiring no extra parameters to be set, i.e., penalty coefficients. This method is used to handle the constraints. Table 4 shows the other GA parameter settings.

Table 4 GA parameter settings

Representation	Real number
Population size	80
Crossover operator	BLX-α [19] (α=0.25)
Crossover probability	1
Mutation operator	[20]
Mutation probability	0.13
Generation gap	95%

The generation gap is the ratio of individuals replaced in each generation. Since this is set to 95%, an elitist GA is used. In addition to elitism, the population size and mutation probability are the most important parameters for the convergence of the GA. There is always a contradiction between computational effort and the risk of premature convergence when using a GA. The population size is here chosen to be 10×8 since the dimensionality of the optimization problem is 8. The mutation probability is set to $1/8$. These values should give a reasonable trade-off between convergence and computational time. It should also be mentioned that a linear ranking scheme is used with selective pressure 1.6.

The computational time required for a single evaluation of the objective function and the constraints is ~5 min. This rather long time depends on the 44 required nonlinear bearing analyses (i.e., four bearings and four cases at 0% load, and seven cases at 100% load). Since a considerable amount of evaluations is necessary in the GA, a parallel implementation is used. The computations are done on a cluster of standard PCs running under Linux with four nodes in the cluster being used in the computations. The results from the optimization are summarized in Sec. 6.

6 Results

The nominal design, given in Table 5 and indicated by the stars in Figs. 4 and 5, gives a total power loss of 523 kW in the bearings. As can be seen from Fig. 4, the temperature constraints for bearings 1 and 2 are not satisfied by the nominal design. The temperatures for bearings 3 and 4 are below the constraint limits. Furthermore, the unbalance constraint cases 2 and 4 (Eq. (5)) are not satisfied by the nominal design. The weak intermediate shaft gives rise to the high unbalance responses in these cases. The nominal design satisfies all the constraints (Eq. (4)) for the different stability cases.

The result from the numerical experiment is shown as dots in Figs. 4 and 5. The dots show the result for 600 randomly generated designs (due to the dense spacing the result almost appears as solid lines in Figs. 4 and 5). Figure 4 shows that it seems unlikely that the bearing temperature constraint is possible to satisfy. Furthermore, none of the 600 randomly generated designs fulfilled the limits for the second and fourth unbalance cases. Hence, finding a feasible solution for the constraints given in Eq. (7) seems very unlikely. Still, it is interesting to know if there exist better designs than the nominal. Therefore, the constraint limits that

Table 5 Nominal and optimal design

Design	B_1 (mm)	ΔR_1 (μm)	B_2 (mm)	ΔR_2 (μm)	B_3 (mm)	ΔR_3 (μm)	B_4 (mm)	ΔR_4 (μm)
Nominal	250	233	250	233	140	291	154	255
Optimal	239	243	241	248	101	256	124	266

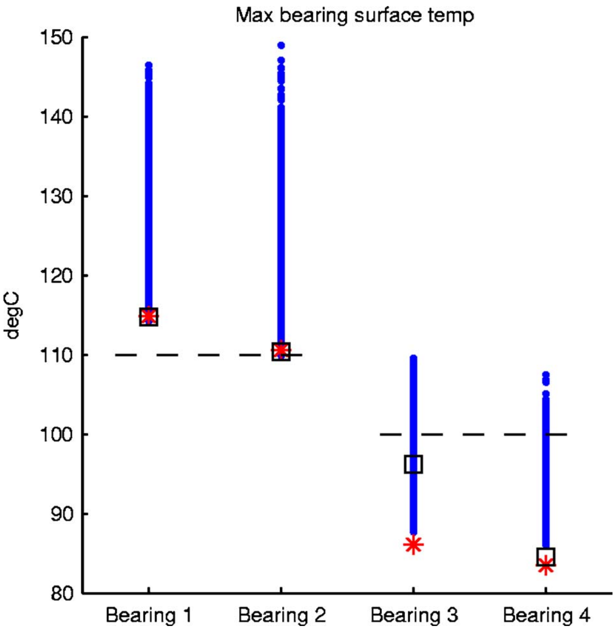


Fig. 4 Bearing temperature for randomly generated designs (dots), nominal design (star), and optimal design (square)

were violated by the nominal design are reformulated and set to the values as in Table 3. Thereby, the existence of at least one feasible design for the reformulated constraints is known. Now it is possible to search for a design with a lower power loss that is not worse in the violated constraints than the nominal design. The search result from the optimization with the reformulated constraint limits is shown in Fig. 6.

Figure 6 shows the mean normalized Euclidian distance in the design variable space between the individuals in the population (right axis). This can be seen as a measure of the convergence in the population. Figure 6 also shows the ratio of feasible solutions (right axis) in the population. This is zero until generation 27 when the first feasible individual appears. The best design (optimal), given in Table 5, is found in generation 96. The power loss for this design is 468 kW. To facilitate comparison, the nominal design is also shown in Table 5. The optimal design is indicated by the squares in Figs. 4 and 5. The widths of the bearings for the optimal design are smaller than for the nominal design. Furthermore, all the radial clearances except for the third bearing are larger for the optimal design than the nominal design. The radial clearances for the first, second, and fourth bearing are the only design variables for the optimal design that are close to the side constraints Eq. (6).

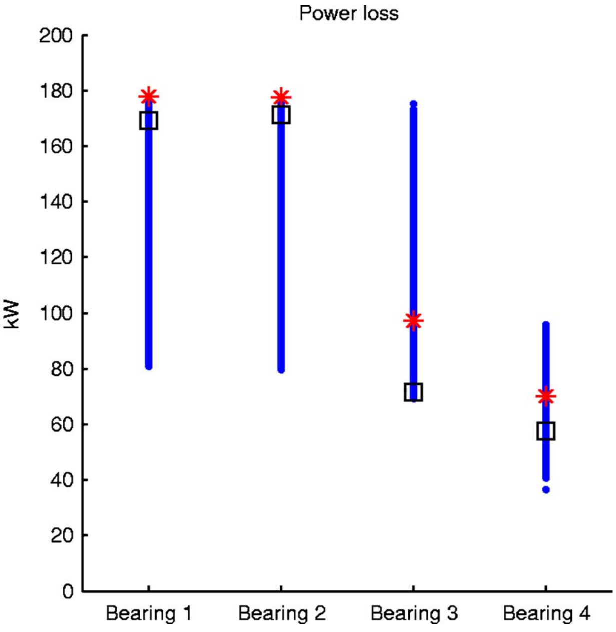


Fig. 5 Power loss for randomly generated designs (dots), nominal design (star), and optimal design (square)

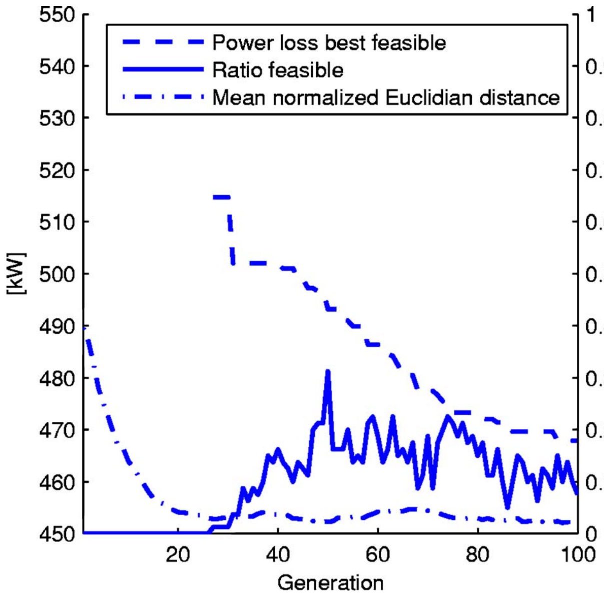


Fig. 6 Search history in optimization

7 Conclusions

In this paper a GA optimization is performed on a gas turbine rotor system with four tilting pad bearings. A real design situation has been formulated as an optimization problem. Nonlinear bearing analyses and several load cases are included. According to the numerical experiment, it is likely impossible to find a solution that satisfies the bearing temperature and unbalance response constraints. Still the optimization shows that it is possible, with minor modifications of the bearings, to find a better design than the nominal design. This design does not violate any of the constraints more than the nominal design and gives 10.5% reduced power loss. Hence, a significant improvement was possible to achieve on a rotor bearing system that is regarded in the industry as well developed. Practically, it would be difficult to find this solution without the use of a search method.

Another conclusion of the work is that the constraint-handling method [8] is able to locate the feasible region even for this highly constrained problem. Hence, it seems to be robust and work well. In this case, only a single optimization run is performed, which shows the strength of the used constraint-handling method. It works in the first shot since no problem-dependent parameters must be specified. Finally, an important aspect that not yet has been addressed is the robustness of the found optimal solution. This is, however, a matter for further research.

Acknowledgment

The first author gratefully acknowledges the involved persons at Demag Delaval Industrial Turbomachinery AB for their support. Furthermore, ALSTOM Power Sweden AB is acknowledged for the funding of the project.

References

[1] Hashimoto, H., and Matsumoto, K., 2001, "Improvement of Operating Characteristics of High-Speed Hydrodynamic Journal Bearings by Optimum Design: Part I—Formulation of Methodology and Its Application to Elliptical Bearing Design," ASME J. Tribol., **123**, pp. 305–312.

[2] Chen, T. Y., and Wang, B. P., 1993, "Optimum Design of Rotor-Bearing Systems With Eigenvalue Constraints," ASME J. Eng. Gas Turbines Power, **115**, pp. 256–260.

[3] Shiau, T. N., and Chang, J. R., 1993, "Multi-Objective Optimization of Rotor-Bearing System With Critical Speed Constraints," ASME J. Eng. Gas Turbines Power, **115**, pp. 246–255.

[4] Choi, B. K., and Yang, B. S., 2001, "Optimal Design of Rotor-Bearing Systems Using Immune-Genetic Algorithm," ASME J. Vibr. Acoust., **123**, pp. 398–401.

[5] Choi, S. P., Kim, Y. C., and Yang, B. S., 2002, "Optimum Design for Rotor-Bearing Systems Using Advanced Genetic Algorithm," *Proceedings of the 9th International Symposium on Transport Phenomena and Dynamics of Rotating Machinery*, Honolulu, Feb. 10–14.

[6] Angantyr, A., and Aidanpää, J.-O., 2004, "Optimization of a Rotor-Bearing System With an Evolutionary Algorithm," *Proceedings of the 10th International Symposium on Transport Phenomena and Dynamics of Rotating Machinery*, D. Bohn ed., Honolulu, March 7–11, RWTH Aachen University, Aachen, pp. 95–96.

[7] Gen, M., and Cheng, R., 2000, *Genetic Algorithms & Engineering Optimization*, Wiley, New York.

[8] Angantyr, A., Andersson, J., and Aidanpää, J.-O., 2003, "Constrained Optimization Based on a Multiobjective Evolutionary Algorithm," *Proceedings of the Congress on Evolutionary Computation*, Canberra, Australia, R. Sarker et al. eds. IEEE, New York, 3, pp. 1560–1567.

[9] Choi, B. K., and Yang, B. S., 2000, "Optimum Shape Design of Rotor Shafts Using Genetic Algorithm," J. Vib. Control, **6**, pp. 207–222.

[10] Choi, B. K., and Yang, B. S., 2001, "Multiobjective Optimum Design of Rotor-Bearing Systems With Dynamic Constraints Using Immune-Genetic Algorithm," ASME J. Eng. Gas Turbines Power, **123**, pp. 78–81.

[11] Vance, J. M., 1988, *Rotordynamics of Turbomachinery*, Wiley, New York.

[12] Genta, G., 1999, *Vibration of Structures and Machines*, Springer-Verlag, Berlin.

[13] Mittwollen, N., and Glienicke, J., 1990, "Operating Conditions of Multi-Lobe Journal Bearings Under High Thermal Loads," ASME J. Tribol., **112**, pp. 330–340.

[14] Cloud, C. H., Foiles, W. C., Li, G., Maslen, E. H., and Barret, L. E., 2002, "Practical Applications of Singular Value Decomposition in Rotordynamics," *Proceedings of the Sixth International Conference on Rotor Dynamics*, E. J. Hahn and R. B. Randall, eds., Sydney, September 30–October 4, UNSW Printing Services, Sydney, Vol. 1, pp. 429–438.

[15] Olausson, H-L., and Klang, A., 1988, "Calculation of Unbalance Sensitivity in Complex Rotor Systems," *Proceedings of Vibrations in Rotating Machinery: International Conference*, Edinburgh, September, IMechE, Edinburgh, pp. 531–537.

[16] Montgomery, D. C., 2001, *Design and Analysis of Experiments*, Wiley, New York.

[17] Kim, H. M., 2001, "Target Cascading in Optimal System Design," Ph.D. thesis, Department of Mechanical Engineering, University of Michigan, Ann Arbor.

[18] Richardson, J. T., Palmer, M. R., Liepins, G., and Hilliard, M., 1989, "Some Guidelines for Genetic Algorithms With Penalty Functions," *Proceedings of the 3rd International Conference on Genetic Algorithms*, J. D. Schaffer, ed., Morgan Kaufmann, Reading, MA, pp. 191–197.

[19] Eshelman, L. J., and Schaffer, J. D., 1993, "Real-Coded Genetic Algorithms and Interval-Schemata," *Foundations of Genetic Algorithms 2*, L. D. Whitley, ed., Morgan Kaufmann, San Mateo, CA, pp. 187–202.

[20] Mühlenbein, H., and Schlierkamp-Voosen, D., 1993, "Predictive Models for the Breeder Genetic Algorithm: I. Continuous Parameter Optimization," Evol. Comput., **1**(1), pp. 25–49.

Nagaraj K. Arakere[1]
e-mail: nagaraj@ufl.edu

Erik Knudsen

Mechanical and Aerospace Engineering,
University of Florida,
Gainesville, FL 32611-6300

Gregory R. Swanson
NASA Marshall Space Center Flight Center,
ED22, Structural Mechanics Group,
Huntsville, AL
e-mail: greg.swanson@nasa.gov

Gregory Duke
JE Sverdrup,
Huntsville, AL
e-mail: greg.duke@msfc.nasa.gov

Gilda Ham-Battista
ERC, Inc.,
Huntsville, AL
e-mail: battista@msfc.nasa.gov

Subsurface Stress Fields in Face-Centered-Cubic Single-Crystal Anisotropic Contacts

Single-crystal superalloy turbine blades used in high-pressure turbomachinery are subject to conditions of high temperature, triaxial steady and alternating stresses, fretting stresses in the blade attachment and damper contact locations, and exposure to high-pressure hydrogen. The blades are also subjected to extreme variations in temperature during start-up and shutdown transients. The most prevalent high-cycle fatigue (HCF) failure modes observed in these blades during operation include crystallographic crack initiation/propagation on octahedral planes and noncrystallographic initiation with crystallographic growth. Numerous cases of crack initiation and crack propagation at the blade leading edge tip, blade attachment regions, and damper contact locations have been documented. Understanding crack initiation/propagation under mixed-mode loading conditions is critical for establishing a systematic procedure for evaluating HCF life of single-crystal turbine blades. This paper presents analytical and numerical techniques for evaluating two- and three-dimensional (3D) subsurface stress fields in anisotropic contacts. The subsurface stress results are required for evaluating contact fatigue life at damper contacts and dovetail attachment regions in single-crystal nickel-base superalloy turbine blades. An analytical procedure is presented for evaluating the subsurface stresses in the elastic half-space, based on the adaptation of a stress function method outlined by Lekhnitskii (1963, Theory of Elasticity of an Anisotropic Elastic Body, Holden-Day, Inc., San Francisco, pp. 1–40). Numerical results are presented for cylindrical and spherical anisotropic contacts, using finite element analysis. Effects of crystal orientation on stress response and fatigue life are examined. Obtaining accurate subsurface stress results for anisotropic single-crystal contact problems require extremely refined 3D finite element grids, especially in the edge of contact region. Obtaining resolved shear stresses on the principal slip planes also involves considerable postprocessing work. For these reasons, it is very advantageous to develop analytical solution schemes for subsurface stresses, whenever possible. [DOI: 10.1115/1.2180276]

Introduction

Single-crystal nickel-base superalloy turbine blades are especially prone to fretting/contact fatigue damage because the subsurface shear stresses induced by fretting action at the damper contact and blade attachment regions can result in crystallographic initiation and crack growth along octahedral planes. The presence of fretting in conjunction with a mean stress in the body of a component can lead to a marked reduction in high-cycle fatigue (HCF) life, sometimes by a factor as great as 10 [1,2]. Fretting occurs when assemblies of components, such as blade and disk attachment surfaces, bolt flanges, snap fit areas, and other clamped members, are subjected to vibration, resulting in contact damage. The combined effects of corrosion, wear, and fatigue phenomena at the fretting contact facilitate the initiation and subsequent growth of cracks.

Currently, the most widely used single-crystal nickel-base turbine blade superalloys are PWA 1480, PWA 1484, RENE' N-5, and CMSX-4. These alloys play an important role in commercial, military and space propulsion systems [3–7]. Military gas turbine mission profiles are characterized by multiple throttle excursions associated with maneuvers, such as climb, intercept, and air-to-air combat. This mission shifts attention to fatigue and fracture considerations associated with areas below the blade platform that contain various stress risers in the form of buttresses and attachments. Blade-disk attachment areas and blade frictional damping devices are particularly prone to fretting/galling fatigue damage [4]. Rocket engine service presents another set of requirements that shifts emphasis to low-temperature fatigue and fracture capability with particular attention given to environmental effects (i.e., high-pressure hydrogen gas exposure, thermal, and cryogenic). Attention has shifted from oxidation erosion, creep, stress rupture, and creep fatigue damage mechanisms to the micromechanics of fatigue and fracture observed between room temperature and 871 °C (1600 °F). Fatigue crack initiation, threshold and region II fatigue crack growth are of primary importance, and the demand for improvements in fracture mechanics properties for turbine blade alloys is imminent [4].

Study of crack initiation under mixed-mode loading is important for understanding fretting fatigue crack initiation in single crystals. The subsurface shear stresses induced by fretting action can result in crystallographic initiation of failure, under mixed-mode loading conditions. Fretting fatigue at low slip amplitudes that induces little or no surface damage can result in greatly reduced fatigue life with accelerated subsurface crystallographic crack initiation, akin to subsurface shear-stress-induced rolling bearing fatigue. The complex interaction between the effects of

---

[1] To whom correspondence should be addressed.

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received August 2, 2004; final manuscript received November 3, 2005. Review conducted by B. Annigeri.

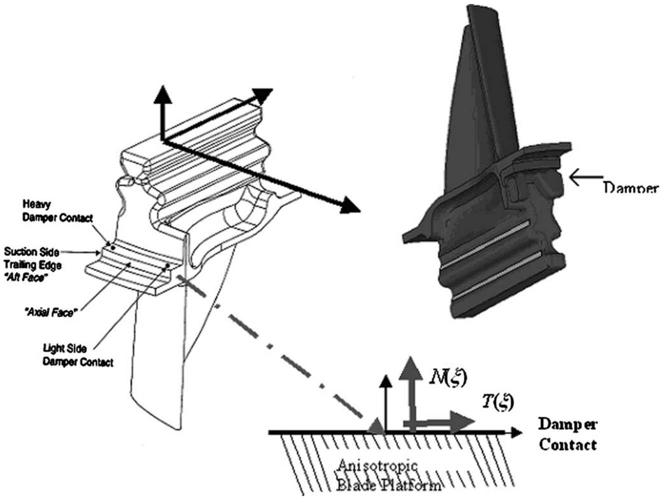


Fig. 1 Damper contact locations on the turbine blade

environment and stress intensity determines which point-source defect species initiates a crystallographic or noncrystallographic fatigue crack [4–7].

This paper presents analytical and numerical methods to evaluate the subsurface stresses in face-centered-cubic (FCC) single-crystal cylindrical and spherical contacts as a function of crystallographic orientation, and contact loads. The subsurface stresses evaluated are subsequently used to assess contact fatigue life, based on a fatigue life model developed previously [8–11]. The motivation for this work is provided by the crystallographic subsurface cracks induced at the damper contact locations in single-crystal turbine blades. Figure 1 shows a schematic of the damper contact location. Figure 2 shows a close-up view of the subsurface induced crystallographic crack propagating on intersecting octahedral planes, ultimately resulting in a pyramidal hole in the blade platform [12].

There is a considerable body of work done on fretting fatigue damage of isotropic polycrystalline materials. Some representative examples are by Hills and Nowell [1], Giannokopoulos et al. [13], Szolwinsky and Farris [14], Attia and Waterhouse [15], Hoeppner [16], Vingsbo and Soderberg [17], and Ruiz et al. [18]. However, studies on subsurface contact stresses and mechanics of fretting fatigue crack initiation and crack growth in orthotropic single-crystal materials are very few. There is an extensive body of literature available in the classical area on the evaluation of subsurface stresses for nonconformal contacts in isotropic materials using analytical methods [19]. However, the amount of published literature involving analytical solutions in anisotropic nonconformal contacts is considerably less. Green and Zerna [20] looked at the two-dimensional (2D) anisotropic contact problem in 1954, for a specific type of anisotropy. Willis [21] examined the Hertzian elliptical contact problem for anisotropic half-spaces using a Fourier transform method. Turner [22] examined the spherical contact between transversely isotropic nonconformal bodies. Fan and Keer [23] examine the 2D contact problem using the analytic function continuation approach based on the Stroh formulation [24]. Vlassak et al. [25] looked at calculating an effective indentation modulus for anisotropic contacts. Analytical solutions were developed for indenters of arbitrary shape being pressed into an anisotropic half-space. For axisymmetric indenters, a limited family of Green's functions is used to obtain a solution for the displacement field. This solution is denoted as an equivalent isotropic solution.

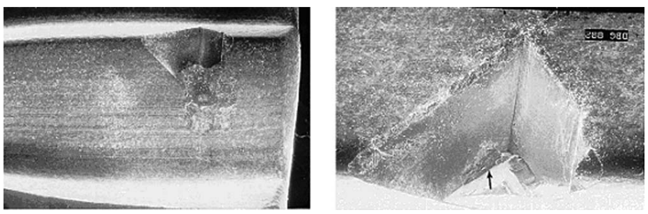


Fig. 2 Crystallographic crack initiation at the damper contact location shown in Fig. 1 [12]

In this paper, we present an analytical solution for a cylindrical contact, using the stress function approach outlined by Lekhnitskii [26], for an anisotropic half-space under conditions of generalized plane strain problem. Finite element subsurface stress results of the cylindrical and spherical anisotropic contacts modeling the damper contact locations shown in Fig. 1 are also presented.

Deformation Mechanisms and Elastic Anisotropy in FCC Single Crystals

Nickel-based single-crystal materials are precipitation strengthened, cast, monograin superalloys based on the Ni-Cr-Al system. The microstructure consists of ≈60–70% by volume of γ' precipitates in a γ matrix. The γ' precipitate, based on the intermetallic compound Ni_3Al, is the strengthening phase in nickel-base superalloys and is a face-centered-cubic (FCC) structure. The γ' precipitate suspended within the γ matrix also has a FCC structure and is comprised of nickel with cobalt, chromium, tungsten, and tantalum in solution [4].

Deformation mechanisms in single crystals are primarily dependent on microstructure, orientation, temperature, and crystal structure. The operation of structures at high temperature places additional materials constraints on the design that are not required for systems that operate at or near room temperatures. In general, materials become weaker with increasing temperature due to thermally activated processes, such as multiple slip and cross-slip. At temperatures in excess of approximately half the homologous temperature (the ratio of the test temperature to the melting point, $=T/T_m$), diffusion controlled processes (e.g., recovery, recrystallization, dislocation climb, and grain growth) become important, which results in further reductions in strength. Slip in metal crystals often occurs on planes of high atomic density in closely packed directions. The four octahedral planes corresponding to the high-density planes in the FCC crystal have three primary slip directions (easy slip) resulting in 12 independent primary $\langle 110\rangle$ $\{111\}$ slip systems. The four octahedral slip planes also have three secondary slip directions resulting in 12 secondary $\langle 112\rangle$ $\{111\}$ slip systems, which represent twinning systems. In addition, the three cube slip planes have two slip directions resulting in six independent $\langle 110\rangle$ $\{100\}$ cube slip systems. Thus, there are 12 primary and 12 secondary slip systems associated with the four octahedral planes and six cube slip systems with the three cube planes, for a total of 30 slip systems [27]. At high temperatures, slip has been observed in non-close-pack directions on the octahedral plane, and on the cube plane, in FCC crystals. The analysis presented in this paper is restricted to the 12 primary $\langle 110\rangle$ $\{111\}$ slip systems only.

Elastic response of FCC crystals is obtained by expressing Hooke's law for materials with cubic symmetry. The generalized Hooke's law for a homogeneous anisotropic body in Cartesian coordinates (x, y, z with origin at point O) is given by Eq. (1) [26,27].

$$\{\varepsilon\}=[\boldsymbol{a}_{ij}]\{\sigma\} \tag{1}$$

$[\boldsymbol{a}_{ij}]$ is the matrix of 36 elastic coefficients, of which only 21 are independent, since $[\boldsymbol{a}_{ij}]=[\boldsymbol{a}_{ji}]$. The elastic properties of FCC crystals exhibit cubic symmetry, also described as cubic syngony. The elastic properties of materials with cubic symmetry can be described with three independent constants designated as the elastic modulus, shear modulus, and Poisson ratio [26], and hence, $[\boldsymbol{a}_{ij}]$ can be expressed as shown in Eq. (2), in the material coordinate

Table 1 Slip plane and slip direction for the 12 primary octahedral slip systems [27]

Slip system	Slip plane $\langle 110\rangle\{111\}$	Slip direction
1	(111)	$[10\bar{1}]$
2	(111)	$[0\bar{1}1]$
3	(111)	$[1\bar{1}0]$
4	$(\bar{1}1\bar{1})$	$[10\bar{1}]$
5	$(\bar{1}1\bar{1})$	$[110]$
6	$(\bar{1}1\bar{1})$	$[011]$
7	$(1\bar{1}\bar{1})$	$[110]$
8	$(1\bar{1}\bar{1})$	$[0\bar{1}1]$
9	$(1\bar{1}\bar{1})$	$[101]$
10	$(\bar{1}\bar{1}1)$	$[011]$
11	$(\bar{1}\bar{1}1)$	$[101]$
12	$(\bar{1}\bar{1}1)$	$[1\bar{1}0]$

system (FCC crystal axes are parallel to x-, y-, and z- coordinate axes). In contrast to the FCC single-crystal material, an isotropic material can only have two independent elastic constants

$$[\boldsymbol{a_{ij}}]=\begin{bmatrix} a_{11} & a_{12} & a_{12} & 0 & 0 & 0 \\ a_{12} & a_{11} & a_{12} & 0 & 0 & 0 \\ a_{12} & a_{12} & a_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & a_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & a_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & a_{44} \end{bmatrix}; \quad \begin{aligned} a_{11} &= \frac{1}{E_{xx}} \\ a_{44} &= \frac{1}{G_{yz}} \\ a_{12} &= -\frac{\nu_{yx}}{E_{xx}} = -\frac{\nu_{xy}}{E_{yy}} \end{aligned} \tag{2}$$

The elastic constants in the generalized Hooke's law of an anisotropic body, $[\boldsymbol{a_{ij}}]$, vary with the direction of the coordinate axes. For orientations other than the (x,y,z) axes, the $[\boldsymbol{a_{ij}}]$ matrix varies with the crystal orientation. In the case of an isotropic body, the constants are *invariant* in any orthogonal coordinate system. Consider a Cartesian coordinate system (x',y',z') that has rotated about the origin O of (x,y,z). The elastic constant matrix $[\boldsymbol{a'_{ij}}]$ in the (x',y',z') coordinate system that relates $\{\varepsilon'\}$ and $\{\sigma'\}$ ($\{\varepsilon\prime\}=[\boldsymbol{a'_{ij}}]\{\sigma\prime\}$) is given by the following transformation [26]:

$$[\boldsymbol{a'_{ij}}]=[\boldsymbol{Q}]^T[\boldsymbol{a_{ij}}][\boldsymbol{Q}]=\sum_{m-1}^{6}\sum_{n-1}^{6} a_{mn}Q_{mi}Q_{nj}, \quad (i,j=1,2,\ldots,6) \tag{3}$$

The transformation matrix $[\boldsymbol{Q}]$ is a 6×6 matrix that is a function of the direction cosines between the (x,y,z) and (x',y',z') coordinate axes. Knowing the state of stress at a given location, in the material coordinate system (x,y,z), the resolved shear stresses on the 12 primary octahedral slip systems, denoted by $\tau^1,\tau^2,\ldots,\tau^{12}$, can be readily obtained using the transformation given by Eq. (4) [28]. The slip plane and slip direction of the 12 primary octahedral slip systems are given in Table 1 [27]. The resolved shear stresses on the secondary octahedral and cube planes are obtained using similar expressions [8,27].

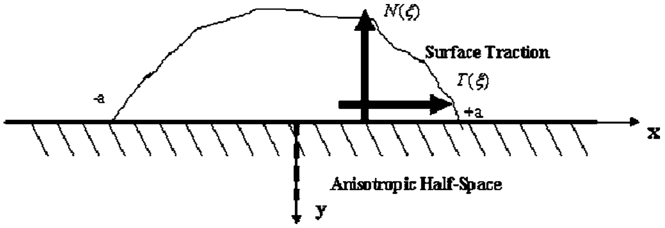


Fig. 3 Anisotropic elastic half-space under generalized plane deformation subjected to normal and tangential traction forces

$$\begin{Bmatrix} \tau^1 \\ \tau^2 \\ \tau^3 \\ \tau^4 \\ \tau^5 \\ \tau^6 \\ \tau^7 \\ \tau^8 \\ \tau^9 \\ \tau^{10} \\ \tau^{11} \\ \tau^{12} \end{Bmatrix} = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 & 0 & -1 & 1 & 0 & -1 \\ 0 & -1 & 1 & -1 & 1 & 0 \\ 1 & -1 & 0 & 0 & 1 & -1 \\ -1 & 0 & 1 & 1 & 0 & -1 \\ -1 & 1 & 0 & 0 & -1 & -1 \\ 0 & 1 & -1 & -1 & -1 & 0 \\ 1 & -1 & 0 & 0 & -1 & -1 \\ 0 & 1 & -1 & -1 & 1 & 0 \\ 1 & 0 & -1 & -1 & 0 & -1 \\ 0 & -1 & 1 & -1 & -1 & 0 \\ -1 & 0 & 1 & -1 & 0 & -1 \\ -1 & 1 & 0 & 0 & 1 & -1 \end{bmatrix} \begin{Bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \sigma_{xy} \\ \sigma_{zx} \\ \sigma_{yz} \end{Bmatrix} \tag{4}$$

Analytical Solution for Two-Dimensional Stress Distribution (Generalized Plane Deformation) in an Anisotropic Elastic Half-Space

The damper contact regions shown in Fig. 1 will be modeled as an elastic anisotropic half-space. This approximation is reasonable since Hertzian-type contact stresses are confined to very small volumes in the vicinity of the contact. An analytical procedure will be presented for evaluating the subsurface stresses in the elastic half-space using a stress-function approach outlined by Lekhnitskii [26]. Lekhnitskii's method for a general anisotropic body has been adapted for a orthotropic FCC single-crystal half-space. Figure 3 shows the elastic half-space subjected to normal traction $N(\xi)$ and tangential traction $T(\xi)$ over the region −a to +a on the x-axis. The traction forces are independent of z, and functions of x and y only. The stresses are also functions of x and y only.

The equilibrium equations under generalized plane strain conditions, for an anisotropic half-space, are expressed as follows [26]:

$$\begin{aligned} \frac{\partial\sigma_x}{\partial x}+\frac{\partial\tau_{xy}}{\partial y}&=0 \\ \frac{\partial\sigma_y}{\partial y}+\frac{\partial\tau_{xy}}{\partial x}&=0 \\ \frac{\partial\tau_{xz}}{\partial x}+\frac{\partial\tau_{yz}}{\partial y}&=0 \end{aligned} \tag{5}$$

Note that the third equilibrium equation in Eqs. (5) is not used for plane strain condition for isotropic materials. However, because of shear coupling induced by anisotropy, the shear stresses τ_{xz} and τ_{yz} are nonzero and functions of x and y.

The stress-strain relations, as defined by the Hooke's law, Eqs. (1), are given by

$$\begin{aligned}
\varepsilon_x &= a_{11}\sigma_x + a_{12}\sigma_y + \cdots + a_{16}\tau_{xy} \\
\varepsilon_y &= a_{12}\sigma_x + a_{22}\sigma_y + \cdots + a_{26}\tau_{xy} \\
&\cdot \\
&\cdot \\
\gamma_{xy} &= a_{16}\sigma_x + a_{26}\sigma_y + \cdots + a_{66}\tau_{xy}
\end{aligned} \tag{6}$$

where $[\boldsymbol{a_{ij}}]$ are a function of crystallographic orientation.

Under the assumptions of generalized plane strain, the subsurface stresses due to the *applied* traction forces can be determined as outlined below.

The stress functions are given by

$$\phi_1'(z) + \phi_2'(z) + \lambda_3\phi_3'(z) = -\frac{1}{2\pi i}\int_{-a}^{+a}\frac{N(\xi)}{\xi - z}d\xi \tag{7}$$

$$\mu_1\phi_1'(z) + \mu_2\phi_2'(z) + \mu_3\lambda_3\phi_3'(z) = -\frac{1}{2\pi i}\int_{-a}^{+a}\frac{T(\xi)}{\xi - z}d\xi \tag{8}$$

$$\lambda_1\phi_1'(z) + \lambda_2\phi_2'(z) + \phi_3'(z) = 0 \tag{9}$$

The μ_i are the roots of the cylindrical characteristic equation, given by Eq. (10), and $z = x + \mu y$.

$$l_4(\mu)l_2(\mu) - l_3^2(\mu) = 0 \tag{10}$$

$$I_2(\mu) = \beta_{55}\mu^2 - 2\beta_{45} + \beta_{44}$$

$$I_3(\mu) = \beta_{15}\mu^3 - (\beta_{14} + \beta_{56})\mu^2 + (\beta_{25} + \beta_{46})\mu - \beta_{24}$$

$$I_4(\mu) = \beta_{11}\mu^4 - 2\beta_{16}\mu^3 + (2\beta_{12} + \beta_{66})\mu^2 - 2\beta_{26}\mu + \beta_{22}$$

$$\lambda_1 = -\frac{l_3(\mu_1)}{l_2(\mu_1)}, \quad \lambda_2 = -\frac{l_3(\mu_2)}{l_2(\mu_2)}, \quad \lambda_3 = -\frac{l_3(\mu_3)}{l_4(\mu_3)}$$

$$\beta_{ij} = a_{ij} - \frac{\alpha_{i3}\alpha_{j3}}{\alpha_{33}} \tag{11}$$

The matrix $\boldsymbol{a_{ij}}$ relates the strains to the stresses. The $\boldsymbol{a_{ij}}$ are functions of the crystal orientation. The stresses are then given by

$$\begin{aligned}
\sigma_x &= 2\,\mathrm{Re}[\mu_1^2\phi_1'(z) + \mu_2^2\phi_2'(z) + \mu_3^2\lambda_3\phi_3'(z)] \\
\sigma_y &= 2\,\mathrm{Re}[\phi_1'(z) + \phi_2'(z) + \lambda_3\phi_3'(z)] \\
\tau_{xy} &= -2\,\mathrm{Re}[\mu_1\phi_1'(z) + \mu_2\phi_2'(z) + \mu_3\lambda_3\phi_3'(z)] \\
\tau_{xz} &= 2\,\mathrm{Re}[\mu_1\lambda_1\phi_1'(z) + \mu_2\lambda_2\phi_2'(z) + \mu_3\phi_3'(z)] \\
\tau_{yz} &= -2\,\mathrm{Re}[\lambda_1\phi_1'(z) + \lambda_2\phi_2'(z) + \phi_3'(z)] \\
\sigma_z &= -\frac{1}{a_{33}}[a_{13}\sigma_x + a_{23}\sigma_y + a_{34}\tau_{yz} + a_{35}\tau_{xz} + a_{36}\tau_{xy}] \\
z_i &= x + \mu_i y
\end{aligned} \tag{12}$$

The normal traction force $N(\xi)$ used is the Hertzian cylindrical contact pressure as $N(\xi) = p_o\sqrt{1 - \xi^2/a^2}$ and $T(\xi) = \mu_f p_o\sqrt{1 - \xi^2/a^2}$, where p_o is the peak pressure and μ_f the coefficient of friction. The tangential traction force is based on a sliding contact and not a contact in partial slip. It must be noted that we are *not* solving a contact problem here, but rather a stress analysis problem in an elastic anisotropic half-space subject to normal and tangential traction forces. The contact dimensions are obtained from a simulated Hertzian cylindrical contact. The semi-elliptical normal pressure distribution seen in isotropic cylindrical contacts is also true for anisotropic contacts [11]. The stress solution has been programed and subsurface stresses computed for various crystal orientations.

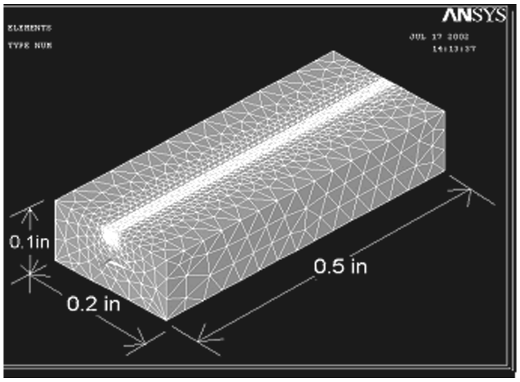


Fig. 4 Three-Dimensional FE model of the elastic anisotropic half space [28]

Figure 4 shows the finite element model used for the numerical results. This ANSYS model represents an elastic anisotropic half-space and was developed using eight-node hexahedral elements (SOLID45s) in the contact region and four-node tetrahedral elements (SOLID45s) in the far field. The load functions, $N(\xi) = p_o\sqrt{1 - \xi^2/a^2}$ and $T(\xi) = \mu_f p_o\sqrt{1 - \xi^2/a^2}$ were applied directly to the finite element model. This applied stress problem does not require the use of contact elements.

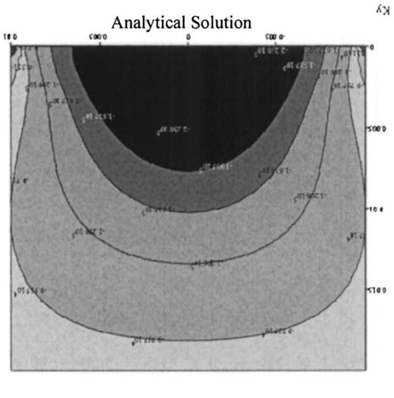
Figure 5 shows a comparison of the analytical and finite element σ_y stress fields for $a = 0.01$ inch, $p_o = 260$ ksi, and for the (x, y, z) axes parallel to the edges of the FCC crystal, i.e., $x = \langle 100\rangle$, $y = \langle 010\rangle$ and $z = \langle 001\rangle$ (Case A). The analytical solution shows excellent agreement with the finite element numerical solution. The FEA solution is evaluated at the midplane, where generalized plane strain conditions prevail. It was observed that the stress field approaches the 2D generalized plane strain solution after a short distance from the edges, indicating that the analytical solution could be used effectively for many practical 3D contact problems. The advantage of this analytical solution is that it is accurate and extremely quick to compute anywhere in the computational domain. The subsurface stress solutions are discussed in greater detail in later sections.

Finite Element Analysis (FEA) of the Cylindrical Anisotropic Contact Problem

A cylindrical indenter on an anisotropic half-space contact model (Fig. 6) was developed in ANSYS. The cylindrical indenter and plate were modeled using eight-node elements (SOLID45s). Surface-to-surface contact elements (CONTA174 and TARGE170) were used at the interface of the two bodies. Because of very high stress gradients in the contact region, a highly refined FE mesh must be used to obtain reliable stress solutions. The densely meshed regions in both the half-cylinder and half-space have roughly the same element size. The refined mesh and the iterative solution of the contact problem require computationally intensive resources, both in time and space. A typical analysis takes two CPU hours on a 2.4 GHz multiprocessor PC-based workstation.

The analytical solution outlined in the previous section can be used for obtaining subsurface stresses in a half-space for a known

Analytical Solution



Finite Element Solution

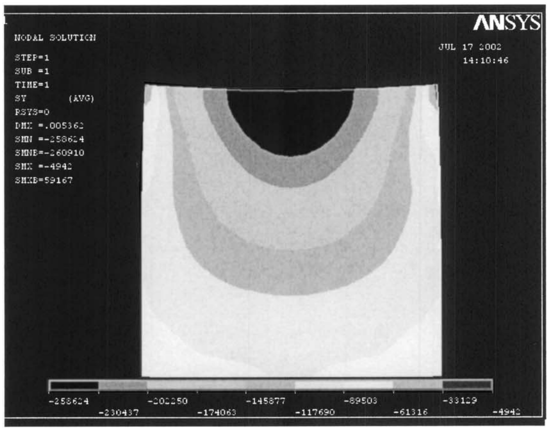


Fig. 5 Stress (σ_y) contours using analytical solution and finite element (ANSYS) solution [28]

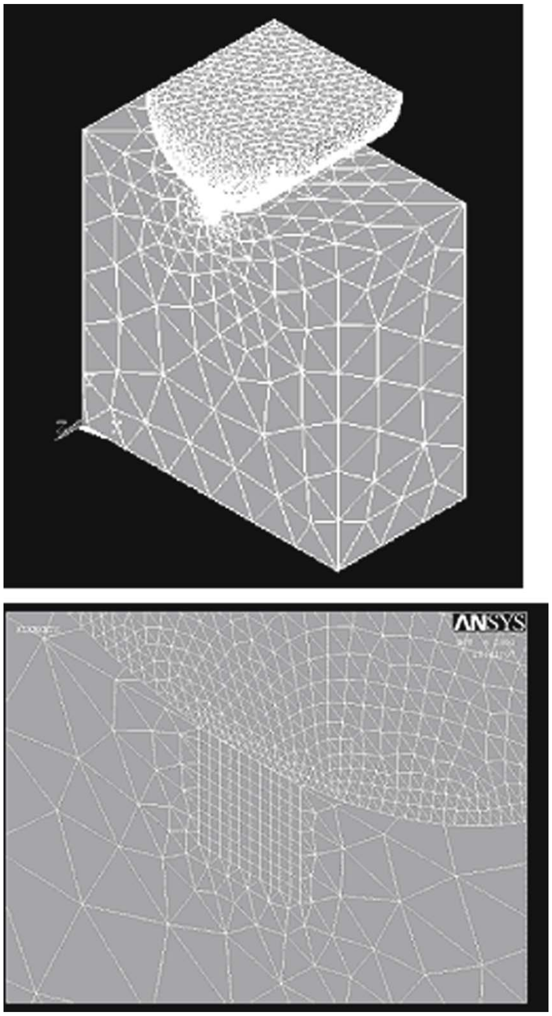


Fig. 6 Three-dimensional FE model of a cylindrical anisotropic contact and close-up view of the meshed contact region [28]

or applied normal and tangential tractions. The analytical solution was obtained by applying a semi-elliptical normal pressure distribution over the contact width. The blades are subjected to fretting stresses at the attaehment regions; however, this analysis (both the analytical and numerical solutions) does not include friction at the contact and, hence, no tangential tractions are generated. The authors are extending this work by examining the effects of friction in a subsequent paper. Fatigue considerations are made toward the end of the paper and the data used in it. An approach for modeling the influence of crystal orientation on fatigue life is given in the same section, but no numerical modeling involving crack initiation and propagation is presented here.

The contact width, $2a$, was estimated initially using a Hertzian isotropic calculation. It should be pointed out that the loads applied here are not the exact same loads imparted on the blades during operation. These loads are, in fact, larger, and this was done to make the contact width large enough so the available resolution (element sizes of 0.001 in.) could be utilized. Thus, the effects of plastic deformation, which could arise from high loads, are not included here. The FEA contact model converges to the correct anisotropic contact width. Figure 7 shows a representative comparison of subsurface stresses computed using the analytical solution and FEA contact model (Fig. 6). Excellent agreement is seen between the two solutions.

The crystallographic orientations are designated by successive rotations about the (XYZ) axes, as follows: γ is rotation about X axis, Δ is rotation about Y' axis, and θ is rotation about Z'' axis. Table 2 shows four different crystallographic orientations considered in the analysis. Figure 8 shows some representative contour plots of the resolved shear stress values for slip systems τ_1, τ_3, and τ_{11}, for two different crystallographic orientations: Case B ($\Delta=15$ deg, $\gamma=0$ deg, $\theta=0$ deg) and case C ($\Delta=-15$ deg, $\gamma=0$ deg, $\theta=0$ deg). The plane on which the crack will nucleate will depend on the magnitude of the shear stress amplitude. Even though the contour plots show similarity between cases B and C, the revolved shear stress (RSS) values are very different because of material orthotropy, and the stresses are a strong function of both primary and secondary crystal orientation.

FEA of the Spherical Anisotropic Contact Problem

A 3D FEA of the spherical anisotropic contact problem was also performed using ANSYS. Figure 9 shows the FEA model of an isotropic spherical contact on a single-crystal plate. The sphere was modeled with eight-node hexahedral elements (SOLID45s), assuming linear-elastic isotropic material behavior. The plate was modeled with 20-node hexahedral elements (SOLID95s) and ten-node tetrahedral elements (SOLID95s). Linear-elastic anisotropic material properties were used in the plate. The contacting surface between the two bodies was represented using ANSYS surface-to-surface contact elements with friction (CONTA174-TARGE170). The indenter or damper is subjected to both normal and tangential loads, and therefore, frictional effects are incorporated.

An analytical solution for the 3D anisotropic contact problem was also obtained using the stress-function approach outlined in Lekhnitskii [26]. However, because of the complexity of the 3D

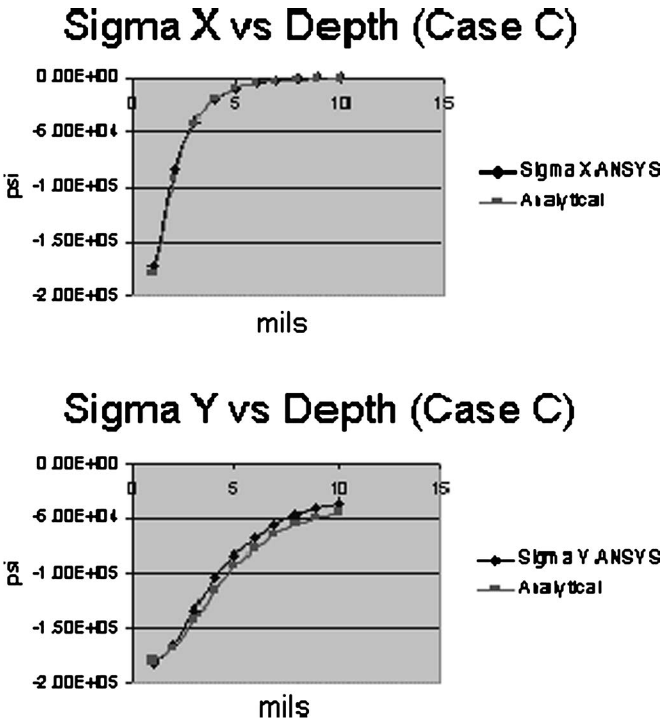


Fig. 7 Comparison of FEA contact and analytical subsurface stresses σ_x and σ_y, as a function of depth, for crystallographic orientation defined by case C ($\Delta=-15$ deg, $\gamma=0$ deg, $\theta=0$ deg) [28]

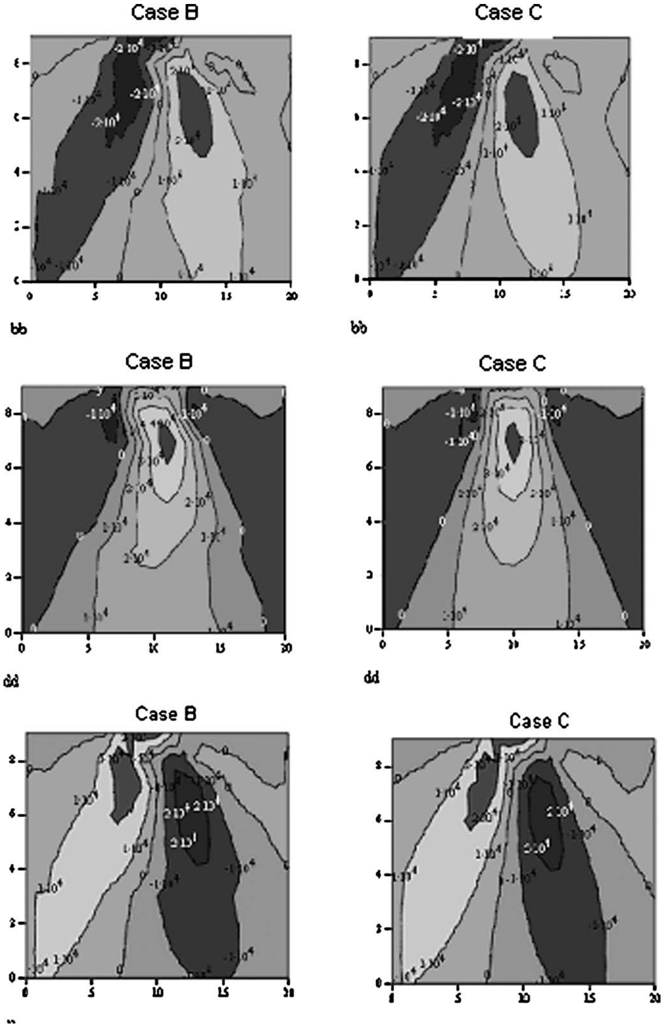


Fig. 8 Contour plots of RSS τ_1, τ_3, and τ_{11}, for cases B and C under the contact region [28]

analytical solution, its presentation is reserved for a separate paper. The numerical results based on the FEA are reported herein.

The Hertzian solution for a spherical isotropic contact on a flat plate is given by

$$P=\frac{4a^3}{3R}(E^*);\quad E^*=\left(\frac{1-\nu_1^2}{E_1}+\frac{1-\nu_2^2}{E_2}\right)^{-1};$$
$$\frac{1}{R}=\frac{1}{R_1}+\frac{1}{R_2};\quad p_o=\frac{1.5P}{\pi a^2} \tag{13}$$

where R_1 and R_2 are the radii, and E_1 and E_2 the Young's moduli of the contacting spheres, respectively, E^* is the effective or composite modulus at the contact, R the composite radius, P the normal load, and a is the contact radius. To derive an effective modulus for the single-crystal orthotropic contact, we refer to Turner's paper [22]. We have adapted his work for a transversely isotropic contact to an orthotropic contact in question. The stress-strain relation in the material coordinate system is given by

$$\begin{bmatrix}\varepsilon_{xx}\\ \varepsilon_{yy}\\ \varepsilon_{zz}\\ \gamma_{xz}\\ \gamma_{yz}\\ \gamma_{xy}\end{bmatrix}=\begin{bmatrix}\frac{1}{E_p} & \frac{-\nu_p}{E_p} & \frac{-\nu_{zp}}{E_{zp}} & 0 & 0 & 0\\ \frac{-\nu_p}{E_p} & \frac{1}{E_p} & \frac{-\nu_{zp}}{E_{zp}} & 0 & 0 & 0\\ \frac{-\nu_{zp}}{E_{zp}} & \frac{-\nu_{zp}}{E_{zp}} & \frac{1}{E_z} & 0 & 0 & 0\\ 0 & 0 & 0 & \frac{1}{2G_{zp}} & 0 & 0\\ 0 & 0 & 0 & 0 & \frac{1}{2G_{zp}} & 0\\ 0 & 0 & 0 & 0 & 0 & \frac{1+\nu_p}{E_p}\end{bmatrix}\begin{bmatrix}\sigma_{xx}\\ \sigma_{yy}\\ \sigma_{zz}\\ \tau_{xz}\\ \tau_{yz}\\ \tau_{xy}\end{bmatrix} \tag{14}$$

The five elastic constants in transverse isotropic constitutive equations are the Young's modulus and Poisson's ratio in the x-y symmetry plane, E_p and ν_p, the Young's modulus and Poisson's ratio in the z direction, E_{pz} and ν_{pz}, and the shear modulus in the z direction G_{zp}. The solution for the orthotropic spherical contact can be derived as

Table 2 Designation of crystallographic orientations

Case	Δ deg	γ deg	θ deg
A	0	0	0
B	+15	0	0
C	−15	0	0
D	0	0	40

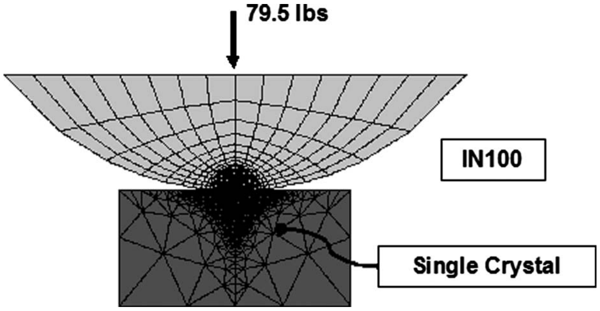


Fig. 9 Three-dimensional FE model of the spherical isotropic indenter on a single-crystal orthotropic substrate

$$P=\frac{4a^3}{3R}\left(\frac{\pi^2\phi_2\kappa}{\phi_1}\right)$$

$$E_{\text{ortho}}=\frac{\pi^2\phi_2\kappa}{\phi_1} \tag{15}$$

where E_{ortho} is an effective contact modulus that can be used to estimate the contact patch size, for the single-crystal contact.

$$\phi_1=\frac{2\gamma}{\sqrt{\alpha}-\gamma},\quad \phi_2=\frac{2}{\varepsilon\pi(\sqrt{\alpha}-\gamma)}$$

$$\alpha=\left(\frac{\lambda-\nu_{zp}^2}{1-\nu_p^2}\right),\quad \lambda=\frac{E_p}{E_z}$$

$$\kappa=\frac{1}{\pi}\ln\left(\frac{\sqrt{\alpha}+\gamma}{\sqrt{\alpha}-\gamma}\right)$$

$$\beta=\frac{(1+\nu_{zp})-\nu_{zp}(1+\nu_p)}{(1-\nu_p^2)}$$

$$\gamma=\left(\frac{2}{\alpha+\beta}\right)^{1/2}\left(\frac{\alpha}{2}-\frac{\nu_{zp}}{2(1-\nu_p)}\right)$$

$$\varepsilon=\left(\frac{\alpha+\beta}{2}\right)^{1/2}\left(\frac{1-\nu_p}{G_p}\right)$$

$$G_p=\frac{E_p}{2(1+\nu_p)} \tag{16}$$

For example, for crystal orientation case A ($\Delta=0$ deg, $\gamma=0$ deg, $\theta=0$ deg), $R=0.25$ in. and $P=79.5$ lb, we can, using Eq. (15), calculate a contact radius, a_{ortho}, as 0.00814 in. The contact radius, a_{FEA}, using FEA is 0.0092 in. The contact radius for the isotropic Hertzian calculation (Eq. (13)) is $a_{\text{iso}}=0.0104$ in (based on $E_1=18.1^6$ psi, $E_2=31.2^6$ psi, $\nu_1=0.3892$, $\nu_2=0.293$). The effective modulus, E_{ortho}, is very useful for calculating the effective contact radius. Once the effective contact radius a_{ortho} is known, we can calculate the maximum contact pressure p_o. The semi-elliptic pressure distribution, $p_o\sqrt{1-r^2}$, can be applied as a normal pressure on the half-space in the FEA, thus effectively decoupling the contact problem with the subsurface stress calculations and, hence, greatly simplifying the numerical problem.

Table 3 shows the contact patch dimensions calculated using FEA contact elements from a spherical indenter model shown in Fig. 9. For crystal orientation cases A and D, even though these two cases represent very different crystal orientations, the contact radius does not vary significantly from $a_{\text{FEA}}=0.0092$ in. However, it must be pointed out that the FEA mesh size in the contact region was 0.001 in.^{2}, and for better resolution the mesh size has to be refined, further highlighting the problems associated with FEA of anisotropic contact problems. Numerical accuracy issues in subsurface stresses as a function of mesh refinement in contact problems is discussed, in detail, by Beisheim and Sinclair [29]. It is very advantageous to calculate the effective contact radius, a_{ortho}, and solve the applied stress problem, rather than resorting to solving the problem using contact elements. This approach is the most effective way to solve contact problems involving single-crystal substrates, especially for design iterations.

Table 3 Spherical orthotropic contact radius as a function of crystal orientation

Orientation	Contact half-width, a_{fea} (in.)
Case A	0.092
Case D	0.092

Representative subsurface stress results, using this decoupling approach and the full contact solution, are shown in Fig. 10. Comparison shows excellent agreement between the two approaches.

Cylindrical and spherical contact simulations were performed for a wide range of crystallographic orientations. Normal contact pressure for these cases was compared to that of case A ($\Delta=0$ deg, $\gamma=0$ deg, $\theta=0$ deg), where the coordinate axes are parallel to the crystal axes, to see the effect of crystal orientation. It was found that even for large orientation deviations from case A, the normal contact pressure and contact patch size did not vary substantially, indicating that the effective contact modulus, E_{ortho}, and contact width (cylindrical) or radius (spherical), a_{ortho}, are relatively insensitive to variations in crystallographic orientation. The E_{ortho} and a_{ortho} values based on case A orientation can be used for nearly *all practical* blade-casting crystallographic orientation deviations from the ideal. This lends further credibility to using the simulated contact model for FEA, which leads to greatly simplified contact subsurface stress analysis. Even though the contact normal pressure does not change substantially, the subsurface stresses are a strong function of orientation. The simulated contact method is very advantageous for performing repeated sub-

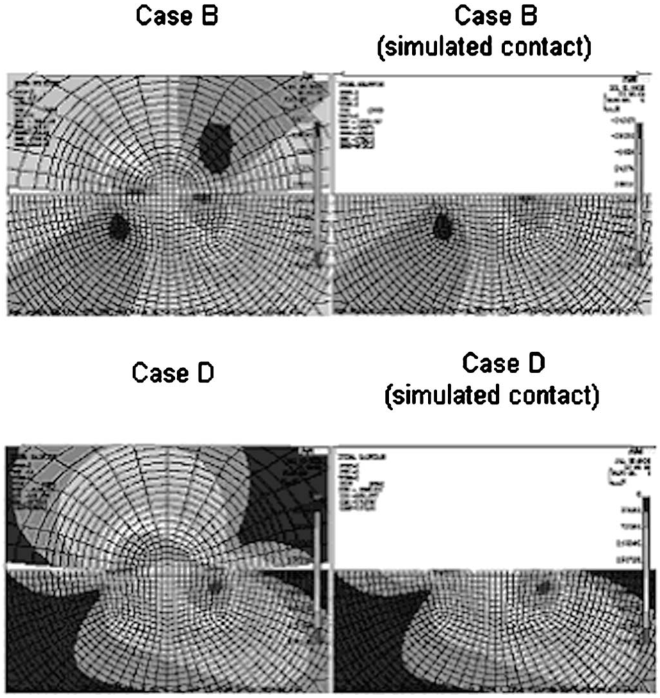


Fig. 10 Comparison of subsurface stresses between the full FEA contact solution and simulated contact, for the orthotropic spherical contact

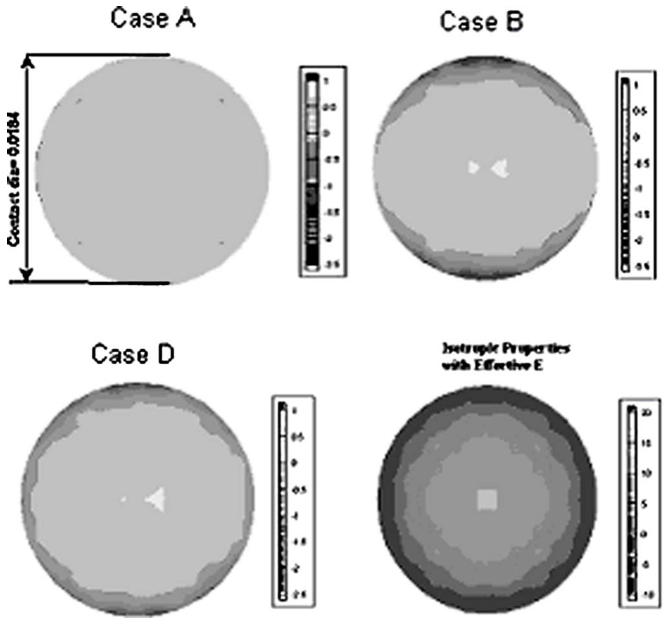


Fig. 11 Weighted percentage difference in normal contact pressure for the orthotropic and isotropic spherical contact, as compared to case A

surface stress calculations required for fatigue life calculations. The full contact solutions typically took 7–8 h to run on a fast multiprocessor PC-based workstation, while the simulated contact took only 1–2 min to execute on the same machine.

Figure 11 shows the weighted percentage differences in contact pressure for some crystal orientations, as compared to case A, for the spherical contact. It can be seen that the deviation of normal pressure from case A is within 2.5% for most practical situations. In contrast, the difference between the isotropic Hertzian contact (Eq. (13)) with case A is significantly higher (within 10%).

Fatigue Considerations

The fatigue crack nucleation and crack growth behavior of single-crystal nickel superalloys is governed by a complex interaction between the operative deformation mechanism, stress intensity, and environmental conditions. The fatigue crack growth behavior is determined by the operative microscopic fracture mode. As a result of the two-phase microstructure present in single-crystal nickel alloys, a complex set of fracture modes exist based on the dislocation motion in the matrix (γ) and precipitate phase (γ'). A fatigue life model was obtained by Swanson and Arakere [8], based on strain-controlled LCF tests conducted at 1200°F in air for single-crystal uniaxial smooth specimens, for four different specimen orientations ⟨001⟩, ⟨111⟩, ⟨213⟩, and ⟨011⟩. Several multiaxial fatigue damage theories, including critical plane methods, were evaluated to identify a suitable fatigue damage parameter that would fit the test data well. The maximum shear stress amplitude, $\Delta\tau_{\max}$, on the slip systems was found to give the best fit for the test data, as shown in Fig. 12. Figure 12 comes from experiments. A power-law curve fit for the data shown in Fig. 12 was used as a fatigue-life estimation equation (1200°F), given below

$$\Delta\tau_{\max} = 397{,}758\ \mathrm{N}^{-0.1598} \tag{17}$$

It should be pointed out that the RSS values on the primary slip systems are calculated based on linear elastic (anisotropic) assumptions and hence nonlinear effects, such as latent hardening, lattice rotation, and twinning in secondary slip systems are not accounted for). Accounting for these effects requires implementation of constitutive relations for crystal plasticity. The fatigue damage parameter $\Delta\tau_{\max}$ has been tested for an extensive set of single-crystal fatigue data, under a range of environmental conditions, and was found to be effective [8–10]. A cylindrical or spherical indenter contacting a single-crystal substrate subject to a vibratory normal and tangential load will result in subsurface cyclic fatigue stresses. These fatigue stresses can lead to subsurface crystallographic cracks, as shown in Fig. 2. Figure 8 shows the contour plots of RSS on the primary octahedral slip systems, for a cylindrical contact loaded with static normal and tangential loads. If the tangential loads are cycled, as would happen during fretting fatigue loading, we can compute the shear stress amplitudes $\Delta\tau_1$, $\Delta\tau_2, \ldots, \Delta\tau_{12}$, in the subsurface region. The subsurface location that yields the maximum $\Delta\tau$ value is likely to initiate a crystallographic fatigue crack.

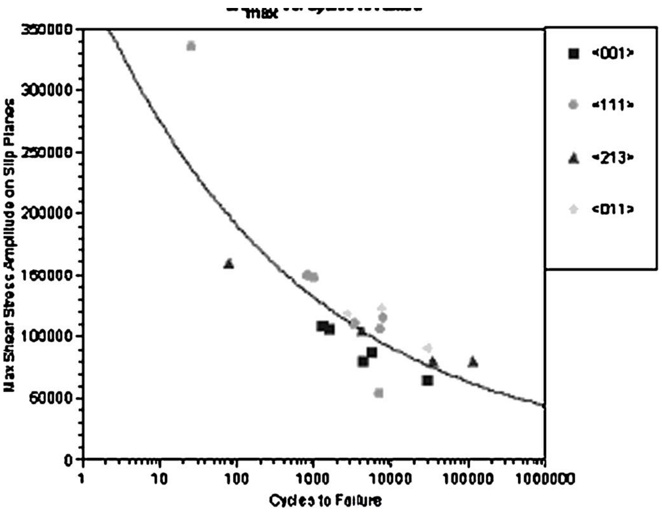


Fig. 12 Fatigue damage parameter, $\Delta\tau_{\max}$ versus cycles to failure [8]

We consider a critical subsurface location near the leading edge of contact, as shown in Fig. 13. We will consider the situation where the tangential traction force $T(x)$ is cycled between a positive and a negative value, and compute the shear stress amplitudes $\Delta\tau$ on the primary planes. Because the secondary crystallographic orientation is not controlled during the blade-casting process, the variation in $\Delta\tau$ due to the variation in secondary orientation alone is of interest. This effect is illustrated in Fig. 14. We see that maximum $\Delta\tau$ ($\Delta\tau_6$ and $\Delta\tau_{11}$, in this case) values vary by 32%

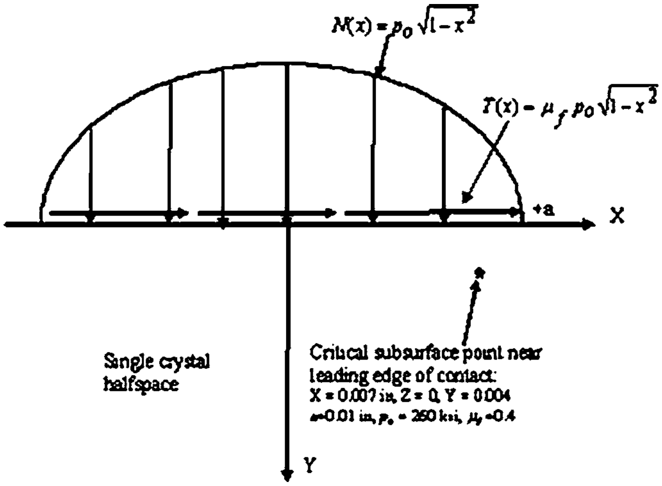


Fig. 13 A critical subsurface point near the leading edge, for a cylindrical single-crystal contact of width 2*a*

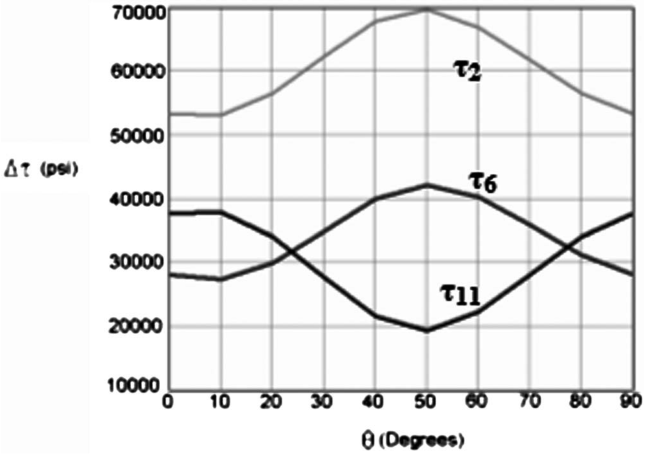


Fig. 14 Variation of $\Delta\tau$ at the critical point shown in Fig. 13 as a function of secondary crystallographic orientation θ (primary orientation=case A)

purely because of variation in secondary crystal orientation between 0 deg and 90 deg. This can result in an order-of-magnitude variation in fatigue life calculated from Fig. 14.

Variation of $\Delta\tau$ is even greater at ~65%, due to a 15 deg primary axis tilt from case A. This can mean *large variations in fatigue life between different blades, under the same loading conditions*, as a result of blade-to-blade variations in primary and secondary crystallographic orientation.

Conclusions

A detailed evaluation of subsurface stresses in cylindrical and spherical orthotropic FCC single-crystal nonconformal contacts is presented, using analytical and numerical techniques. Effects of variation in primary and secondary crystallographic orientation on subsurface stresses are included. Evaluation of subsurface stresses is an essential part of contact fatigue-life calculations at damper contacts and dovetail attachment regions. A two-dimensional analytical solution for subsurface stresses in cylindrical single-crystal contacts is presented, based on an adaptation of a stress function approach by Lekhnitskii [26]. Lekhnitskii's method for an anisotropic half-space in generalized plane deformation has been adapted to a FCC orthotropic half-space. The analytical solution showed excellent agreement with the 3D FEA results. It was observed that the 3D FEA stress field approaches the 2D generalized plane strain solution after a short distance from the edges, indicating that the analytical solution could be used effectively for many practical 3D contact problems. The advantage of the analytical solution is that it is accurate and extremely quick to compute, anywhere in the computational domain.

Three-dimensional FEA results for the spherical single-crystal contact are presented. An effective contact modulus for the single-crystal half-space, E_{ortho}, and contact radius a_{ortho} (Eq. (15)) is shown to be effective in calculating the contact patch size. The FEA of the contact problem can be greatly simplified by using a_{ortho} and applying the normal pressure based on Hertzian assumptions over the contact patch. It is demonstrated that this applied stress problem yields accurate subsurface stresses and greatly simplifies the FEA by avoiding the use contact elements. For a fixed normal load, the E_{ortho} and a_{ortho} values were found to be relatively insensitive to variations in crystallographic orientation. Hence the E_{ortho} and a_{ortho} values based on case A ($\theta=0$) orientation can be used for nearly *all practical* blade-casting crystallographic orientation deviations from the ideal. This lends further credibility to using the simulated contact model for FEA. The simulated contact approach is very advantageous for performing repeated subsurface stress calculations required for fatigue-life evaluation.

It must be noted that even though the contact area and normal pressure do not vary substantially with crystal orientation, the subsurface stresses are a strong function of orientation. Therefore, the resolved shear stresses on the slip systems, and hence, fatigue life, are a strong function of crystallographic orientation. It is shown that there can be an order-of-magnitude variation in contact fatigue life between different blades under the same loading conditions, as a result of blade-to-blade variations in primary and secondary crystallographic orientation.

Obtaining accurate subsurface stress results for anisotropic single-crystal contact problems requires extremely refined 3D finite element grids, especially in the edge of the contact region. Obtaining resolved shear stresses on principal slip planes also involves considerable postprocessing work. For these reasons it is very advantageous to develop analytical solution schemes for subsurface stresses, whenever possible.

Acknowledgment

The authors would like to acknowledge the NASA/ASEE Summer Faculty Fellowship Program and the Graduate Student Accompanying Program. The support from this program, administered by the University of Alabama in Tuscaloosa, enabled Nagaraj K. Arakere and Erik C. Knudsen to work during the Summer of 2002 at the NASA Marshall Space Flight Center, Huntsville, AL, where a majority of this work was completed.

References

[1] Hills, D. A., and Nowell, D., 1994, *Mechanics of Fretting Fatigue*, Kluwer, Dordrecht.
[2] Dombromirski, J., 1990, "Variables of Fretting Process: Are There 50 of them?" *Standardization of Fretting Fatigue Test Methods and Equipment*, ASTM, Philadelphia, pp. 60–68.
[3] Cowles, B. A., 1996, "High Cycle Fatigue in Aircraft Gas Turbines: An Industry Perspective," Int. J. Fract., **80**, pp. 1–16.
[4] Deluca, D., and Annis, C., 1995, "Fatigue in Single Crystal Nickel Superalloys," Office of Naval Research, Department of the Navy, Report No. FR23800, August.
[5] Sims, C. T., 1987, "Superalloys: Genesis and Character," *Superalloys—II*, C. T. Sims, N. S. Stoloff, and W. C. Hagel, eds., Wiley, New York, p. 1.
[6] VerSnyder, F. L., and Guard, R. W., 1960, "Directional Grain Structure for High Temperature Strength," Trans. Am. Soc. Met., **52**, p. 485.
[7] Gell, M., and Duhl, D. N., 1986, "The Development of Single Crystal Superalloy Turbine Blades," *Processing and Properties of Advanced High-Temperature Materials*, S. M. Allen, R. M. Pelloux, and R. Widmer, eds., ASM, Metals Park, OH, p. 41.
[8] Arakere, N. K., and Swanson, G., 2002, "Effect of Crystal Orientation on Fatigue Failure of Single Crystal Nickel Base Turbine Blade Superalloys," ASME J. Eng. Gas Turbines Power, **124**, pp. 161–176.
[9] Swanson, G., and Arakere, N. K., 2000, "Fatigue Failure of Single Crystal Nickel Base Turbine Blade Superalloys," NASA/TP-2000–210074.
[10] Arakere, N. K., 2000, "High Temperature Fatigue Properties of Single Crystal Superalloys in Air and Hydrogen," ASME Paper No. 01-GT-585.
[11] Arakere, N. K., and Swanson, G., 2001, "Analysis of Fretting Stresses in Single Crystal Ni-Base Turbine Blade Attachment Regions," ASME J. Tribol., **123**, pp. 413–423.
[12] DeLuca, D. P., 2000, Pratt & Whitney, East Hartford, CT, personal communication.
[13] Giannokopoulos, A. E., Lindley, T. C., and Suresh, S., 1998, "Aspects of Equivalence Between Contact Mechanics and Fracture Mechanics: Theoretical Connections and a Life-Prediction Methodology for Fretting-Fatigue," Acta Mater., **46**(9), pp. 2955–2968.
[14] Szolwinski, M. P., and Farris, T. N., 1996, "Mechanics of Fretting Fatigue Crack Formation," Wear, **198**, pp. 93–107.
[15] Attia, M. H., and Waterhouse, R. B., eds, 1992, *Standardization of Fretting Fatigue Test Methods and Equipment*, ASTM, Philadelphia.
[16] Hoeppner, D. W., 1990, *Mechanisms of Fretting Fatigue and Their Impact on Test Methods Development, Standardization of Fretting Fatigue Test Methods and Equipment*, ASTM, Philadelphia, pp. 23–32.
[17] Vingsbo, O., and Soderberg, D., 1988, "On Fretting Maps," Wear, **126**, pp. 131–147.
[18] Ruiz, C., Boddington, P. H. B., and Chen, K. C., 1984. "An Investigation of Fatigue and Fretting in a Dovetail Joint," Exp. Mech., **24**(3), pp. 208–217.
[19] Johnson, K. L., 1985, *Contact Mechanics*, Cambridge University Press, Cambridge, England, pp. 84–106.
[20] Green, A. E., and Zerna, W., 1954, *Theoretical Elasticity*, Clarendon Press, Oxford.

[21] Willis, J. R., 1966, “Hertzian Contact of Anisotropic Bodies,” J. Mech. Phys. Solids, **14**, pp. 163–176.
[22] Turner, J. R., 1979, “Contact on a Transversely Isotropic Half-Space, or Between Two Transversely Isotropic Bodies,” Int. J. Solids Struct., **16**, 409.
[23] Fan, H., and Keer, L. M., 1994, “Two-Dimensional Contact on an Anisotropic Half-Space,” ASME J. Appl. Mech., **61**, pp. 250–255.
[24] Stroh, A. N., 1958, “Dislocation and Cracks in Anisotropic Elasticity,” Philos. Mag., **3**, pp. 625–646.
[25] Vlassak, J. J., et al., 2003, “The Indentation Modulus of Elastically Anisotropic Materials for Indenters of Arbitrary Shape,” J. Mech. Phys. Solids, **51**, pp. 1701–1721.
[26] Lekhnitskii, S. G., 1963, *Theory of Elasticity of an Anisotropic Elastic Body*, Holden-Day, San Francisco, pp. 1–40.
[27] Stouffer, D., and Dame, L., 1996, *Inelastic Deformation of Metals: Models, Mechanical Properties, and Metallurgy*, Wiley, New York, pp. 387–417.
[28] Knudsen, E. C., 2003, “Analytical and Numerical Evaluation of Subsurface Stresses in Anisotropic (Single-Crystal) Contacts,” M. S. thesis, Department of Mechanical & Aerospace Engineering, University of Florida, Gainesville.
[29] Beisheim, J. R., and Sinclair, G. B., 2002, “Three-Dimensional Finite Element Analysis of Dovetail Attachments,” ASME Paper No. GT-2002–30306.

A Probabilistic Micromechanical Code for Predicting Fatigue Life Variability: Model Development and Application

K. S. Chan
ASME Fellow

M. P. Enright

Southwest Research Institute ®,
6220 Culebra Road,
San Antonio, TX 78238

This paper summarizes the development of a probabilistic micromechanical code for treating fatigue life variability resulting from material variations. Dubbed MICROFAVA *(micromechanical fatigue variability), the code is based on a set of physics-based fatigue models that predict fatigue crack initiation life, fatigue crack growth life, fatigue limit, fatigue crack growth threshold, crack size at initiation, and fracture toughness. Using microstructure information as material input, the code is capable of predicting the average behavior and the confidence limits of the crack initiation and crack growth lives of structural alloys under LCF or HCF loading. This paper presents a summary of the development of the code and highlights applications of the model to predicting the effects of microstructure on the fatigue crack growth response and life variability of the* $\alpha+\beta$ *Ti-alloy Ti-6Al-4V.* [DOI: 10.1115/1.2180811]

Introduction

Fatigue crack growth properties of structural alloys generally exhibit variability because of variations in microstructure. This variability is traditionally characterized by extensive testing to develop a large database so that the mean, variance, and lower bound of the fatigue properties can be ascertained. An obvious drawback of this approach is that an experimental database is time consuming and costly to generate. Another limitation is that the empirical approach seldom provides information regarding the sensitivities of various material parameters to the probability of fracture associated with a particular component or the means for reducing fatigue variability. One of the contributing factors to this problem is that most existing fatigue life or crack growth models are not linked with the fundamental material parameters; instead, they are based on empirical constants obtained from curve fits of data from physical tests. Although these models characterize the main descriptors (mean, variance) of fatigue life as a function of applied stress range (S-N_f curves) or fatigue crack growth rate, da/dN, as a function of stress intensity factor range ΔK, they provide little or no physical insight into the material parameters that influence failure associated with fatigue crack nucleation and growth.

In general, fatigue life can generally be considered to consist of two parts, which are crack initiation and crack propagation lives. The crack propagation life N_f can be predicted by integrating the da/dN equation, leading to

$$N_f = \int_{a_i}^{a_f} \left[\frac{da}{dN} \right]^{-1} da \tag{1}$$

where a_i and a_f are the initial and final crack length, respectively, and a_f is defined as the critical crack length at fracture. The fatigue crack growth (FCG) life N_f is most sensitive to the initial crack size, a_i [1], but also depends on microstructure, which can alter intrinsic material properties as well as extrinsic properties originating from crack shielding mechanisms.

Three analytical tools are required to address the effects of microstructure variation on fatigue life variability: (i) a microstructure-based crack initiation model for predicting initiation life and initial crack size distribution, (ii) a microstructure-based fatigue crack growth model, and (iii) a probabilistic framework for treating random material variables in the fatigue models. Previously, Tanaka and Mura [2] proposed a microstructure-based model for predicting crack formation at slip bands, inclusions, and pores. Recently, Chan [3] extended Tanaka and Mura's model to predict crack size at initiation on the basis of microstructure parameters, such as grain size, slip plane spacing, and the Taylor factor. Furthermore, a microstructure-based fatigue crack growth model has been developed by Chan and Torng [1], Chan [4], and Chan and Enright [5,6] to provide a relationship between material parameters and fatigue crack growth in the power-law regime that can be used to establish a link between materials design and structural/mechanical component design. The model relates material parameters (e.g., dislocation cell size, dislocation barrier spacing, yield stress, fatigue ductility, Young's modulus) to fatigue crack growth rate in a power-law relation. This model has been subsequently extended and is expressed as a closed-form sum of terms representing all three stages (stage I, II, and III) of fatigue crack resistance [5,6].

For treating fatigue life variability, a probabilistic model [7] has been developed that quantifies crack growth rate variability as a function of stress intensity factor range and randomness of material parameters. The microstructure-based fatigue crack growth model can be reformulated in terms of crack growth life and combined with additional random variables (e.g., defect size, applied stress, inspection intervals, etc.) for application to probabilistic life prediction. The microstructure-based fatigue crack initiation model, fatigue crack growth rate model, fatigue limit model, FCG threshold models, and crack closure models were integrated into a probabilistic fatigue life model named the micromechanical fatigue variability (MICROFAVA), model which is a stand-alone computer code that is executed via a graphical-user interface (GUI) for input, output, and graphical presentation of computational results. The main functions of MICROFAVA are the computation of stress-life ($S-N_f$) curves, da/dN curves, and initial crack size distribution based on the probability distribution of the microstructure size and other material parameters, such as Young's modulus, yield stress, fatigue ductility, and Taylor factor (texture). The out-

---

Contributed by the International Gas Turbine Institute of ASME for publication in the Journal of Engineering for Gas Turbines and Power. Manuscript received August 30, 2005; final manuscript received September 6, 2005. Review conducted by K. C. Hall. Paper presented at the ASME Turbo Expo 2005: Land, Sea, and Air, Reno, NV, Paper No. GT2005-68983.

put of MICROFAVA is designed to provide the relevant fatigue properties and the material data input file for the Southwest Research Institute® (SwRI)-developed probabilistic design and life-prediction code called DARWIN® [8].

The objectives of this paper are (i) to summarize the development of probabilistic fatigue models in MICROFAVA and (ii) to demonstrate the predictive capability of MICROFAVA for probabilistic life prediction of structural alloy Ti-6Al-4V. First, a brief summary of the probabilistic fatigue crack initiation and growth models in MICROFAVA is presented. Second, the application of the probabilistic model to predicting the FCG variability in Ti-6Al-4V is highlighted. Finally, application of the model to lifing a Ti rotor design is demonstrated.

Probabilistic Fatigue Crack Growth Rate Modeling

The physics-based crack growth model proposed by Chan and Enright for treating the three stages of fatigue crack growth has the form given by [5–7]

$$\frac{da}{dN} = \frac{2s\xi^{n_2/2}}{[E(2s)^{1/2}]^{n_2}}\{\Delta K_T^{n_1-n_2}\Delta K_{\text{eff}}^{-n_1} + \Delta K_{\text{eff}}^{-n_2} - [(1-R)K_C]^{-n_2}\}^{-1} \tag{2}$$

with

$$\xi = \frac{Es}{4\sigma_y'\epsilon_f' d} \tag{3}$$

$$d = d_o\left(\frac{D}{D_o}\right)^{\gamma} \tag{4}$$

and

$$\Delta K_T = \frac{\Delta\sigma_y'}{\Delta\sigma_e}\Delta K_{th} \tag{5}$$

where the fatigue crack growth rate, da/dN, is expressed in terms of an effective stress intensity range ΔK_{eff}, which is the difference between the applied stress intensity range ΔK and the stress intensity range at crack closure ΔK_{cl}; ξ, a dimensionless normalizing parameter defined in terms of the Young's modulus (E), dislocation cell size (s), cyclic yield stress (σ_y'), fatigue ductility coefficient (ϵ_f'), and the dislocation barrier spacing (d). The dislocation barrier spacing is taken to be a function of the grain size D as described by Eq. (4), where d_o, D_o, and γ are material constants.

The stress intensity range at the stage I to stage II transition ΔK_T, is a function of the fatigue limit ($\Delta\sigma_e$), cyclic yield stress range and the large-crack FCG threshold (ΔK_{th}); n_1 and n_2 are the stage I and stage II exponents, respectively. R is the stress ratio and K_C is the fracture toughness.

The effective stress intensity range ΔK_{eff} is expressed as [9]

$$\Delta K_{\text{eff}} = U_P U_D U_R U_O \Delta K \tag{6}$$

where U_i is ratio of ΔK_{eff} to ΔK for individual crack closure mechanisms including plasticity- (P), deflection- (D), roughness- (R), and oxide-induced (O) closure. Analytical expressions for computing individual U_i terms, which are summarized in a recent paper [9], include Newman's formulation for treating plasticity-induced crack closure [10], Suresh's formulation for crack deflection and roughness-induced crack closure [11], as well as an oxide-wedging model.

Probabilistic Crack Initiation Modeling

The crack initiation model of Tanaka and Mura [2] was developed based on a dislocation dipole mechanism operating in a surface grain. During fatigue loading, irreversible slip occurs on parallel slip planes in a favorably oriented surface grain, producing dislocation dipoles at the ends of a double pileup whose coalescence ultimately leads to crack nucleation. Chan [3] recently extended Tanaka and Mura's model [2] to explicitly incorporate the crack size, as well as other pertinent material parameters in the response equation by considering the energetics of the fatigue crack initiation process. Specifically, the length of the incipient crack was obtained by equating the elastic strain energy released by dislocation coalescence and crack opening to the fracture energy, consisting of elastic and plastic components, required to form the crack surfaces. This formulation leads one to [3,6]

$$(\Delta\sigma - 2Mk)N_i^{\alpha} = \left[\frac{8M^2\mu^2}{\lambda\pi(1-\nu)}\right]^{1/2}\left(\frac{h}{D}\right)\left(\frac{a}{D}\right)^{1/2} \tag{7}$$

for crack initiation at slip bands, where $\Delta\sigma$ is the stress range, M is the Taylor factor, k is the friction stress, N_i is the cycles-to-initiation, ν is Poisson's ratio, μ is shear modulus, and $\lambda(=0.005)$ is a universal constant. Equation (7) relates the crack initiation life N_i to the grain size D, the slipband width h, and the crack half-length or depth a. The exponent to N_i has been generalized to α where $0 < \alpha \leqslant 1$. The value of α is not a constant, but depends on the degree of slip irreversibility and the stacking fault energy. Expressions for crack initiation at inclusions and notches have also been developed and reported earlier [3].

Probabilistic Framework

For probabilistic modeling, the material-specific variables in Eqs. (2) and (7) are treated separately from the remaining variables. Previous studies [1,7] suggest that some of the material-specific variables can be approximated as deterministic variables (e.g., E, s, K_c). The remaining material variables (dislocation barrier spacing d, cyclic yield stress σ_y', fatigue ductility coefficient ϵ_f') are modeled as random variables. In many instances, these material variables are all functions of the grain size D. If the non-material-specific random variables $(a_i, \Delta K, R)$ are temporarily treated as deterministic variables, Eq. (1) can be used to predict N_f scatter attributed to material variability. A general probabilistic framework has been developed to address the influences of both the material- and non-material-specific variables on the probability of fatigue fracture [12–15]. The N_f scatter results shown in this paper can be directly applied to this framework, allowing the designer to quantify the influences of the fundamental micromechanics variables on overall component risk. Figure 1 shows the schematics of the probabilistic framework utilized in MICROFAVA for modeling random variables.

As in probabilistic crack growth modeling, material variables in Eq. (7) are treated separately from the remaining (e.g., loading) variables in probabilistic crack initiation modeling. The material specific variables include grain size (D), slipband width (h), fatigue limit (Mk), elastic properties, and Taylor factor, which is influenced by texture. The remaining variables are the stress range, cycles-to-crack initiation (N_i), and crack size (a_i) at initiation. Equation (7) can be utilized in two ways: (i) for computing the N_i distribution for a given value of crack size at initiation, a_i, and (ii) for computing the distribution of a_i for a given number of fatigue cycles. For both cases, the general probabilistic framework described by Enright et al. [15] is employed.

Model Application to Ti-6Al-4V

The probabilistic microstructure-based fatigue crack initiation and growth modeling methodology is illustrated for Ti-6Al-4V. Figure 2 shows the typical microstructure of the Ti-6Al-4V alloy, which is comprised of primary α grains and Widmanstatten $\alpha + \beta$ colonies. The volume fraction of α grain was ~60%, and the average grain size was ~12 μm; both were determined using metallographic techniques. In grain size measurement, both the α grain size and the $\alpha+\beta$ colony size were treated as equivalent and fitted to one single probability density function. A log-normal distribution was used to describe the grain size distribution (D) of

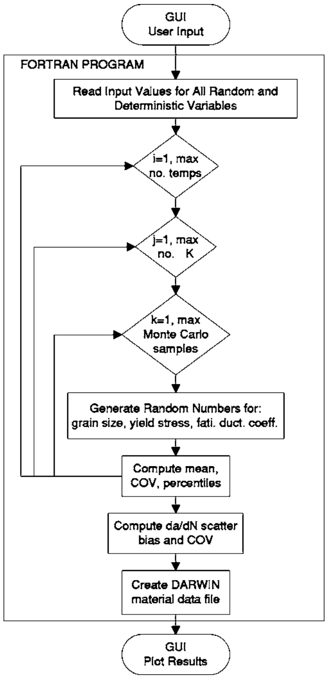


Fig. 1 Schematics of the probabilistic framework in MICROFAVA for modeling material-specific random variables

Ti-6Al-4V. As shown in Fig. 3, this distribution provides conservative values of D compared to experimental values.

Material constants in the FCG model were determined from FCG data for $R=0.8$ to ensure that the FCG response was free from crack closure effects. Values for the material-related variables are indicated in Table 1. Once intrinsic material properties were obtained, the FCG model was utilized to predict deterministic da/dN curves for various R ratios by applying the crack closure model and using average values of grain size, yield stress, and fatigue ductility coefficient. The crack closure model [9] includes plasticity [10], deflection-, and roughness-induced [11] closure mechanisms, but not oxide-induced crack closure. For crack growth in Ti-6Al-4V, grain size was the dominant microstructural variable and was modeled as a log-normal random variable (mean=11.7 μm, standard deviation=3.2 μm) based on grain size measurements shown in Fig. 3. Cyclic yield stress σ_y' and fatigue ductility coefficient ϵ_f' were modeled as deterministic variables to emphasize the influence of grain size variability on da/dN and N_f scatter results.

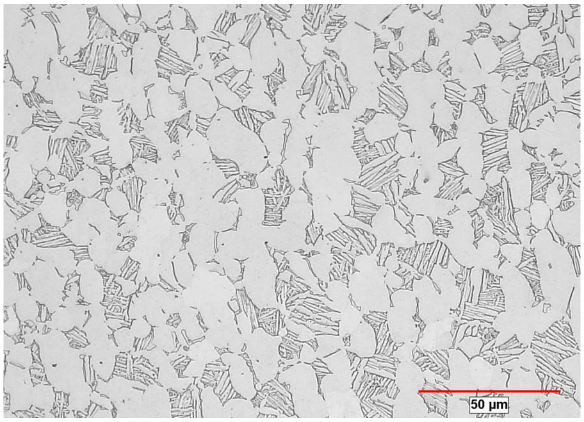


Fig. 2 Microstructure of Ti-6Al-4V shows 60% primary α grain (light phase) and a 40% of $\alpha+\beta$ Widmanstatten colonies

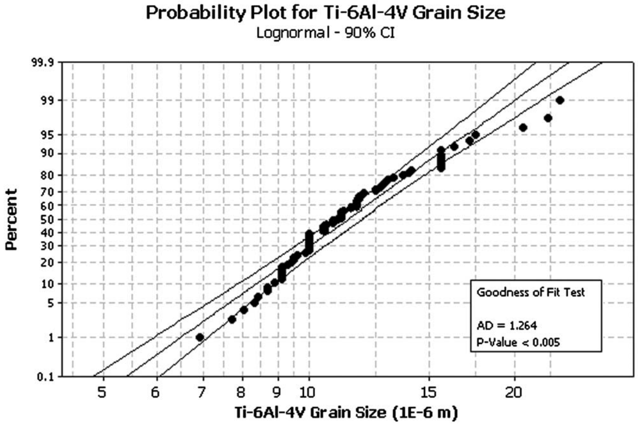


Fig. 3 Grain size of Ti-6Al-4V compared to the log normal distribution

***S-N_f* Variability.** The microstructure-based fatigue crack initiation and growth models were used to predict the fatigue cycles for initiating a thumbnail crack of length $2a$ and depth c and the number of cycles for this crack to reach a critical depth. Values of material constant in the crack initiation model are presented in Table 2.The computed crack initiation life N_i and crack growth life N_g for various alternating stresses σ_a are compared against experimental data in Fig. 4. The corresponding crack length versus fatigue cycle curves of crack initiation (N_i) and crack growth (N_g) for a stress range of 552 MPa at a stress ratio R of 0.1 are

Table 1 Values of deterministic material variables for Ti-6Al-4V at 24 °C used in microstructure-based fatigue crack growth model (Eq. (2))

Variables	Description	Value
E	Young's modulus	1.61×10^5 MPa
n_1	Stage I exponent	23.0
n_2	Stage II exponent	3.87
s	Dislocation cell size	0.25 μm
d_o	Reference dislocation barrier spacing	1.0 μm
D_o	Reference grain size	1.0 μm
R	Stress ratio	0.1
K_c	Fracture toughness	66.7 MPa $\sqrt{m}$
ΔK_{th}	Fatigue threshold	2.0 MPa $\sqrt{m}$
D	Grain size	11.7 μm
σN_y	Cyclic yield stress	909.33 MPa
ϵN_f	Fatigue ductility coefficient	0.0389
γ	Dislocation barrier spacing exponent	1.0

Table 2 Values of deterministic material variables for Ti-Al-4V at 24°C used in microstructure-based fatigue crack initiation model (Eq. (7))

Variables	Description	Value
M	Taylor factor	2
μ	Shear modulus	4.4×10^{-4} MPa
ν	Poisson's ratio	0.333
λ	Universal constant	0.005
α	Fatigue imitation life exponent	0.5
σ_e	Fatigue limit (Mk)	272.5 MPa
h	Slipband width	5×10^{-8} m

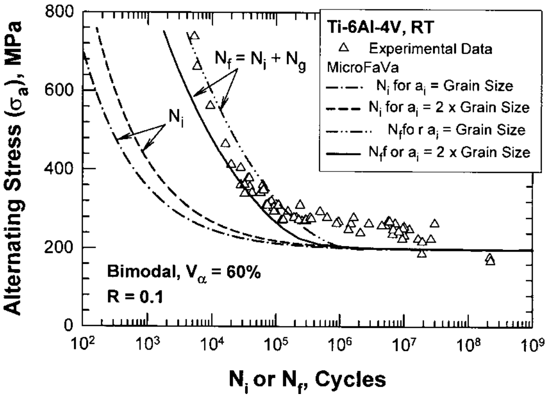


Fig. 4 Predicted crack initiation life, N_i, crack growth life, N_g, and total life, N_f, compared against observed failure life for Ti-6Al-4V. From Chan et al. [16]. Experimental data are from [17].

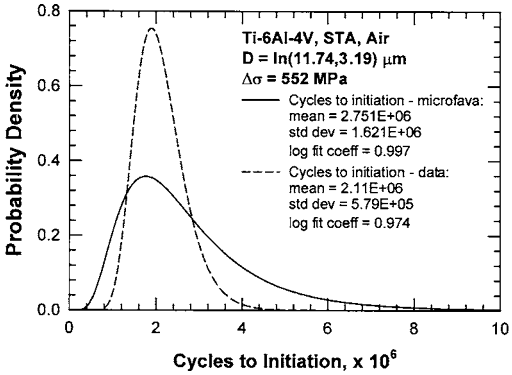


Fig. 6 Measured PDF for cycles-to-initiation compared against MICROFAVA model prediction based on measured grain-size distribution. From Chan et al. [16].

presented in Fig. 5. For both cases, the fatigue crack growth lives N_g were computed without a large-crack threshold. Variations of the N_g value were obtained when a fatigue crack growth threshold was employed. The life variation depends on the crack size at which a crack is considered to have "initiated." Different curves of crack length versus fatigue cycle were predicted for crack initiation and growth. For most cases, the transition from crack initiation to crack growth appeared to occur at a crack size of one to two grain diameters, as shown in Fig. 4 and 5.

The microstructure-based fatigue crack initiation and growth models were utilized to predict the variability of fatigue crack initiation and growth live due to grain size variations. A comparison of the predicted and observed probability density functions (PDF) for crack initiation life is shown in Fig. 6, which shows that the range of initiation life values associated with the experimental data is within the values predicted by MICROFAVA.

Fatigue Threshold Variability. The FCG threshold model in MICROFAVA contains an intrinsic term, which originates from the crack-tip cyclic slip process, and an extrinsic term, which arises from a combination of plasticity-induced closure with crack deflection, asperity-induced, and oxide-induced crack closure mechanisms. In particular, the ΔK_{th} expression is given by [9]

$$\Delta K_{th} = U^{-1}\Delta K_{th,in} = [U_P U_D U_R U_O]^{-1}\Delta K_{th,in} \tag{8}$$

with the intrinsic threshold $\Delta K_{th,in}$ given by [9]

$$\Delta K_{th,in} = 0.1\left[\frac{ME}{\sigma_{ys}}\right]^{1/2} Eb^{1/2} \tag{9}$$

where b is the magnitude of the Burgers vector. The proposed model was applied to compute the ΔK_{th} for Ti-6Al-4V for R values ranging from -1 to 1 and the results are presented in Fig. 7. For these calculations, $\Delta K_{th,in}$ ranges from 1.2 MPa$\sqrt{\text{m}}$ to 2.0 MPa$\sqrt{\text{m}}$. Most of the crack closure was contributed by crack-wake plasticity. Figure 7 shows a comparison of model calculations against experimental data of Ti-6Al-4V from the HCF program [18] at Air Force Research Laboratory (AFRL), Marci [19], Döker and Marci [20], and Boyce and Ritchie [21].

The Ti-6Al-4V alloy studied by Boyce and Ritchie [21] were identical to those used in the HCF program; hence, there was good agreement between these two sets of ΔK_{th} data. For these two sets of data, $\Delta K_{th,in}$=2 MPa$\sqrt{\text{m}}$, it was correctly predicted by the intrinsic threshold model, Eq. (9). The increase in ΔK_{th} at $R \leq 0.8$ was caused entirely by crack closure mechanisms. The lower $\Delta K_{th,in}$ value (1.2 MPa$\sqrt{\text{m}}$) for the Ti-6Al-4V material studied by Marci [19] and Döker and Marci [20] was predicted from Eq. (9) using a Taylor factor of $M \approx 2$. This value of M appeared to give the lower bound for $\Delta K_{th,in}$. Again, the increase in ΔK_{th} at $R < 0.5$ was entirely due to crack closure mechanisms. Thus, the variation of $\Delta K_{th,in}$ from 1.2 MPa$\sqrt{\text{m}}$ to 2 MPa$\sqrt{\text{m}}$ could arise from microstructural effects such as texture (different values of M), yield stress, and cell formation. On the other hand, the variation of ΔK_{th} at low R ratio ($R < 0.5$) is likely caused by crack closure.

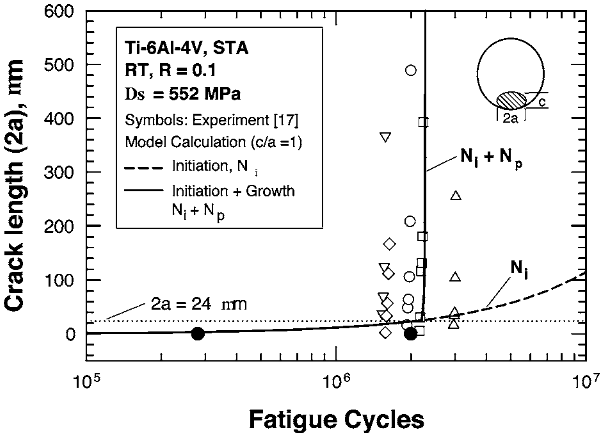


Fig. 5 A comparison of predicted and measured crack lengths versus fatigue cycles for the initiation (N_i) of one grain-sized crack and its growth (N_g) to failure. Experimental data are from [17]. From Chan et al. [16].

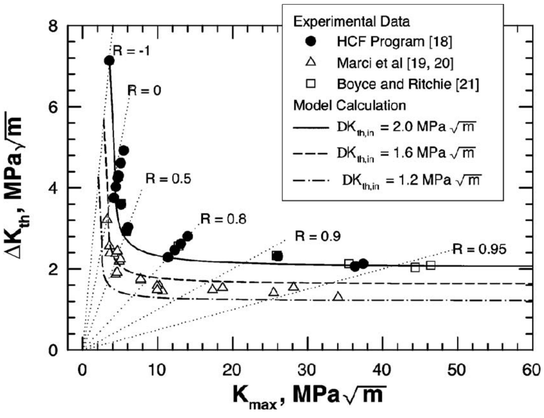


Fig. 7 Comparison of model predictions against measured ΔK_{th} for Ti-6Al-4V as a function of K_{max} From Chan [9]

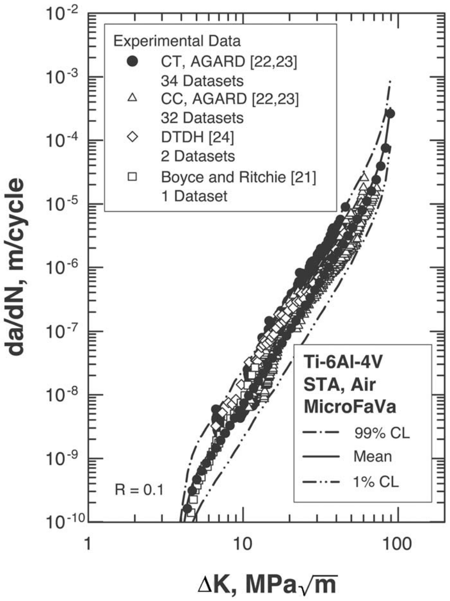


Fig. 8 Measured da/dN data compared against model predictions based on a log-normal grain size distribution [16]

Crack Growth Rate Variability. Crack growth rate (mean and 1%/99% confidence limit) values for $R=0.1$ predicted using Eq. (2) are compared to experimental data [21–24] in Fig. 8. A total of 67 sets of da/dN data of Ti- 6Al-4V of the same or equivalent material and microstructure are presented in Fig. 8. Experimental da/dN data at selected ΔK values were also analyzed to obtain probability densities. These results are compared to probability densities and coefficient of variation (COV) of predicted crack growth rate at selected ΔK values in Figs. 9 and 10. It can be observed from Fig. 9 that the PDF values based on the experimental data are bound by the microstructure-based PDF values for most of the ΔK values considered. Figure 10 shows that the predicted COV is about 0.54–0.56 over the ΔK range of 3–90 $\mathrm{MPa}\sqrt{\mathrm{m}}$. The two largest values of the experimentally determined COV are 0.45 at $\Delta K=15\ \mathrm{MPa}\sqrt{\mathrm{m}}$ and 0.48 at $\Delta K=45\ \mathrm{MPa}\sqrt{\mathrm{m}}$. The larger discrepancy at ΔK values in the range of 18–30 $\mathrm{MPa}\sqrt{\mathrm{m}}$ is due to insufficient experimental data in that crack growth regime.

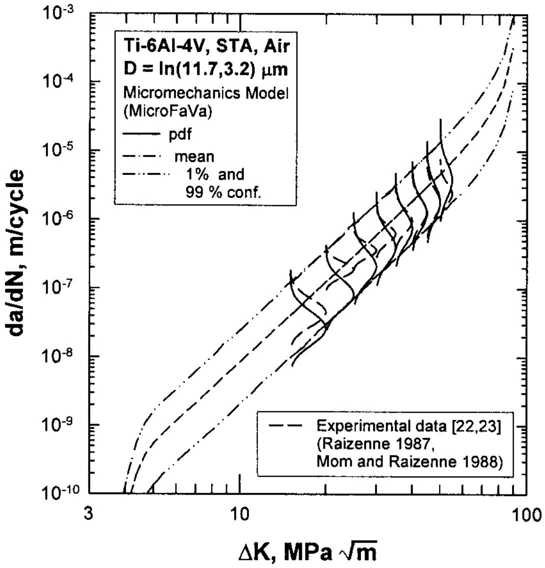


Fig. 9 Comparison of computed crack growth rate predictions and probability density functions with experimental data for $R=0.1$ at 24°C

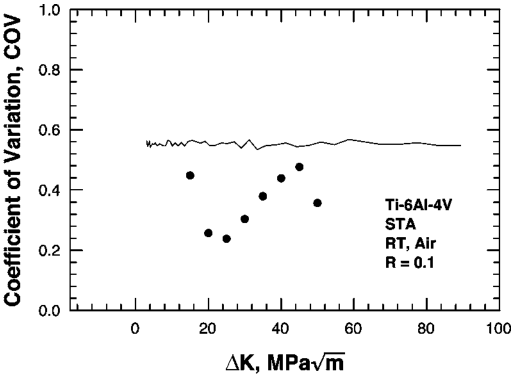


Fig. 10 Comparison of computed crack growth rate (da/dN) coefficient of variation with experimental data for $R=0.1$ at 24°C

Crack Growth Life Variability. Computational and experimental da/dN values were applied to the life prediction of a 10 cm × 10 cm Ti-6Al-4V plate with material properties indicated in Table 1. A uniform stress range of 600 MPa was applied to the plate at a stress ratio R of 0.1 and an initial crack (0.035 cm × 0.035 cm) was placed at one of the corners. Crack growth life was computed using probabilistic fracture mechanics software (DARWIN® 2004 [8]) using: (i) tabular da/dN values based on Eq. (2), and (ii) da/dN slope and intercept constants based on linear and bilinear curve fits of experimental data from the literature [22,23]. The results are shown in Fig. 11. It can be observed that all of the predicted crack growth life values based on the experimental da/dN data fall within the 5% and 95% confidence limits based on life values associated with the computational model (Eq. (2)). The corresponding computed and observed probability density of crack propagation life are compared in Fig. 12, which shows that the predicted mean is in good agreement with the experimental data. On the other hand, the predicted variability is larger than experimental observations. As shown in Fig. 12, the observed COV is ~0.23 while the predicted COV is 0.57. The observed COV is lower partly because of the sampling size and partly because the COVs associated with curve fitting of individual da/dN datasets have not been included in the computation of the experimental COV. The observed COV is expected to ap-

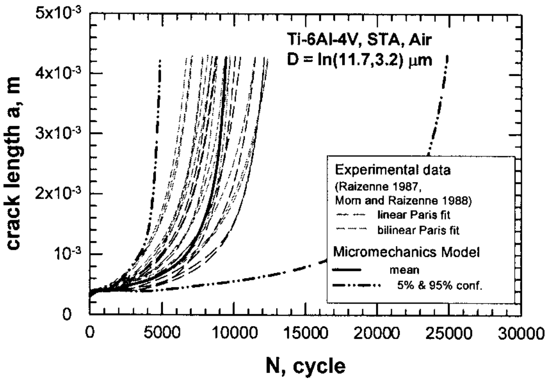


Fig. 11 Comparison of predicted crack growth life values based on microstructure-based computational model and experimental data da/dN results [22,23] for $R=0.1$ at 24°C [16]

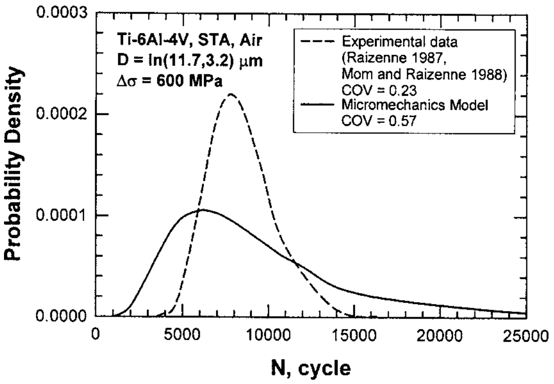


Fig. 12 Comparison of computed and observed probability density of FCG lives based on microstructure-based computational model and experimental data *da/dN* results [22,23] for $R=0.1$ at 24°C

proach the predicted COV as the sampling size increases and the COVs of individual *da/dN* datasets are taken into account. Additional FCG life computations for $\Delta\sigma=552$ MPa indicates that the predicted COV is independent of the applied stress and depends only on the grain size distribution since stress variations are not considered in these calculations.

Application to Probabilistic Life Prediction of Titanium Components

Two example problems were solved to validate the MICROFAVA prediction and its integration with DARWIN. For the first validation case, the calibration problem of a fictitious titanium rotor disk subjected to a single inertial load in FAA Advisory Circular AC33.14-1 [25] and an empirically derived hard α distribution [13] as the initial crack size distribution were used to assess the MICROFAVA/DARWIN *da/dN* and life predictions. Essentially, identical results were obtained by the current methodology and those reported in the FAA Circular. In the second validation, MICROFAVA and DARWIN were used to analyze the probability of fatigue fracture of a representative Ti rotor design. As in the first problem, the initial crack size was described by the distribution of the hard α particle size and a power-law FCG relation was used. Figure 13 shows the rotor disk design divided into 221 zones after five zone refinements. The predicted probability of fracture computed on the basis of conventional (experimental *da/dN* data) material input and MICROFAVA input are compared in Fig. 14.

It should be noted that the results based on the conventional data input did not consider the coefficient of variation (COV) of *da/dN*. Without considering the COV, the DARWIN prediction using MICROFAVA input is slightly less conservative than that using the conventional input. The ability of MICROFAVA to predict the COV in addition to the mean *da/dN* allows one to quantify the increase in the probability of fracture due to *da/dN* variation. As shown in Fig. 14, the predicted probability of fracture is increased slightly above the conventional method line when the COV predicted from MICROFAVA is incorporated into the DARWIN design analysis. These computations illustrate that MICROFAVA is a viable methodology for predicting fatigue variability due to microstructural variations and is highly compatible with probabilistic design and life-prediction methods.

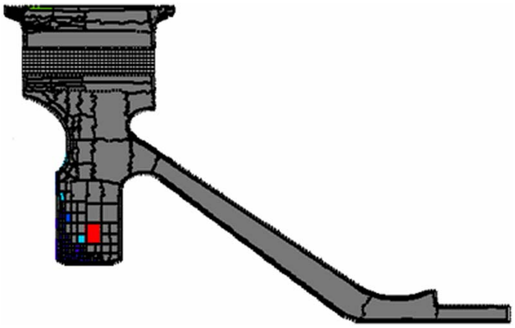


Fig. 13 Analysis of a Ti rotor design using MICROFAVA and DARWIN

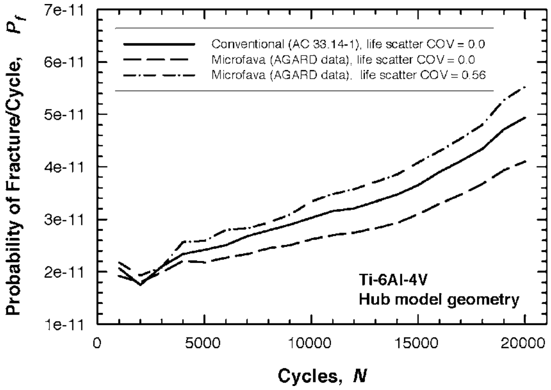


Fig. 14 Computed probability of fracture using MICROFAVA and DARWIN [16]

Conclusions

In this paper, the development of a probabilistic microstructure-based crack initiation and growth model was summarized and the model capability was illustrated for life prediction of Ti-6Al-4V. It was shown that the microstructure-based $S\text{-}N_f$ and *da/dN* simulation results are in agreement with experimental data for the range of values considered in this study. Fatigue life variability in Ti-6Al-4V appears to originate, at least partly if not totally, from microstructural variability in the form grain size variations. In addition, the microstructure-based models appear to provide reasonable estimates of crack growth life and total life variability when compared to life values based on experimental *da/dN* and $S\text{-}N_f$ data. Application of the probabilistic fatigue model to a Ti alloy rotor design illustrates that the proposed model is capable of predicting the confidence bounds of the fatigue life and can significantly reduce the database requirement used to define fatigue life variability.

This work was sponsored by the Air Force Office of Scientific Research (AFOSR), USAF, under Contract No. F49620-01-1-0547, and was performed as part of the AFOSR MEANS program, Dr. Craig S. Hartley, program manager. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research or the U.S. Government.

References

[1] Chan, K. S., and Torng, T.-Y., 1996, "A Probabilistic Treatment of Microstructural Effects on Fatigue Crack Growth of Large Cracks," ASME J. Eng. Mater. Technol., **118**, pp. 379–386.
[2] Tanaka, K., and Mura, T., 1981, "A Dislocation Model for Fatigue Crack Initiation," ASME J. Appl. Mech., **48**, pp. 97–103.
[3] Chan, K. S., 2003, "A Microstructure-Based Fatigue Crack Initiation Model," Metall. Mater. Trans. A, **34**, pp. 43–58.
[4] Chan, K. S., 1993, "Scaling Laws for Fatigue Crack Growth of Large Cracks in Steels," Metall. Trans. A, **24**, pp. 2473–2486.
[5] Chan, K. S., and Enright, M. P., 2002, "Physically-based Models for Predicting Fatigue Life Variability in Ni-Based Superalloys," *Proceedings of Modeling the Performance of Engineering Structure Materials III*, T. Srivatsan, D. Lesuer, and E. Taleff, eds., TMS, Warrendale, PA, pp. 135–142.
[6] Chan, K. S., and Enright, M. P., 2003, "Application of Microstructure-based

Fatigue Models to Component Life Prediction," *Fatigue 2003*, M. R. Bache et al., eds., Engineering Integrity Society, Sheffield, UK, pp. 39–48.

[7] Enright, M. P., and Chan, K. S., 2004, "Extension of a Microstructure-based Fatigue Crack Growth Model for Predicting Fatigue Life Variability," J. ASTM Int., **1**(8), Paper No. JAI11566.

[8] SwRI, 2004, DARWIN®, Design Assessment of Reliability with INspection®, Software Package, Ver. 5.1, Southwest Research Institute, San Antonio.

[9] Chan, K. S., 2004, "Variability of Large-Crack Fatigue Crack Growth Thresholds in Structural Alloys," Metall. Mater. Trans. A, **35**, pp. 3721–3735.

[10] Newman, J. C., Jr., 1984, "A Crack Opening Stress Equation for Fatigue Crack Growth," Int. J. Fract., **24**, pp. R131–R135.

[11] Suresh, S., 1985, "Fatigue Crack Deflection and Fracture Surface Contact: Micromechanical Models," Metall. Trans. A, **16**, pp. 249–260.

[12] Enright, M. P., and Wu, Y. T., 2003, "Probabilistic Fatigue Life Sensitivity Analysis of Titanium Rotors," *Proceedings of the 41st AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Non-Deterministic Approaches Forum*, Atlanta, April 3–6, AIAA Paper No. 2000-1647.

[13] Leverant, G. R., 2000, "Turbine Rotor Material Design - Final Report," DOT/FAA/AR-00/64, Federal Aviation Administration, Washington, DC.

[14] Wu, Y. T., Enright, M. P., and Millwater, H. R., 2002, "Probabilistic Methods for Design Assessment of Reliability With Inspection," AIAA J., **40**(5), pp. 937–946.

[15] Enright, M. P., Huyse, L., McClung, R. C., and Millwater, H. R., 2004, "Probabilistic Methodology for Life Prediction of Aircraft Turbine Rotors," *Proceedings, 9th Biennial ASCE Aerospace Division International Conference on Engineering, Construction and Operations in Challenging Environments - Earth & Space*, R. B. Malla and A. Maji, eds., Houston, March 7–10, ASCE, Reston, VA, pp. 453–460.

[16] Chan, K. S., Enright, M. P., and Kung, J. S., 2005, "MicroFaVa: A Micromechanical Code for Predicting Fatigue Life Variability," *Materials Damage Prognosis*, J. M. Larsen, L. Christodoulou, J. R. Calcaterra, M. L. Dent, M. M. Derriso, W. J. Hardman, J. W. Jones, and S. M. Russ, eds., TMS, Warrendale, PA, pp. 135–142.

[17] Bellows, R. S., Muju, S., and Nicholas, T., 1999, "Validation of the Step Test Method for Generating Haigh Diagrams for Ti-6Al-4V," Int. J. Fatigue, **21**, pp. 687–697.

[18] Sheldon, J. W., Bain, K. R., and Donald, J. K., 1999, "Investigation of the Effects of Shed-Rate, Initial K_{max}, and Geometric Constraint on ΔK_{th} in Ti-6Al-4V at Room Temperature," Int. J. Fatigue, **21**, pp. 733–741.

[19] Marci, G., 1994, "Comparison of Fatigue Crack Propagation Thresholds of Two Ti Turbine-Disk Materials," Int. J. Fatigue, **16**, pp. 409–412.

[20] Döker, H., and Marci, G., 1983, "Threshold Range and Opening Stress Intensity Factor in Fatigue," Int. J. Fatigue, **5**, pp. 187–191.

[21] Boyce, B. L., and Ritchie, R. O., 2001, "Effect of Load Ratio and Maximum Stress Intensity on the Fatigue Threshold in Ti-6Al-4V," Eng. Fract. Mech., **68**, pp. 129–147.

[22] Raizenne, M. D., 1987, "Fatigue Crack Growth Rate Data for AGARD TX114 Engine Disc Cooperative Test Programme," Laboratory Memorandum ST-479, National Aeronautical Establishment, Ottowa, Canada, September.

[23] Mom, A. J. A., and Raizenne, M. D., 1988, "AGARD Engine Disc Cooperative Test Programme," AGARD Report No. 766, August.

[24] 1994, *Damage Tolerant Design Handbook*, Purdue University, West Lafayette, Vol. 3, WL-TR-94-4043.

[25] U.S. Department of Transportation, Federal Aviation Administration, 2001, "Advisory Circular—Damage Tolerance for High Energy Turbine Engine Rotors," AC33.14-1, Washington, DC, January.

Rotordynamic Coefficients Measurements Versus Predictions for a High-Speed Flexure-Pivot Tilting-Pad Bearing (Load-Between-Pad Configuration)

Adnan M. Al-Ghasem

Dara W. Childs

Texas A&M University,
Turbomachinery Laboratory,
College Station, TX 77840

Experimental dynamic force coefficients are presented for a four pad flexure-pivot tilting-pad bearing in load-between-pad configuration for a range of rotor speeds and bearing unit loadings. Measured dynamic coefficients have been compared to theoretical predictions using an isothermal analysis for a bulk-flow Navier-Stokes (NS) model. Predictions from two models—the Reynolds equation and a bulk-flow NS equation models are compared to experimental, complex dynamic stiffness coefficients (direct and cross-coupled) and show the following results: (i) The real part of the direct dynamic-stiffness coefficients is strongly frequency dependent because of pad inertia, support flexibility, and the effect of fluid inertia. This frequency dependency can be accurately modeled for by adding a direct added-mass term to the conventional stiffness/damping matrix model. (ii) Both models underpredict the identified added-mass coefficient ($\sim$32 kg)*, but the bulk-flow NS equation predictions are modestly closer. (iii) The imaginary part of the direct dynamic-stiffness coefficient (leading to direct damping) is a largely linear function of excitation frequency, leading to a constant (frequency-independent) direct damping model. (iv) The real part of the cross-coupled dynamic-stiffness coefficients shows larger destabilizing forces than predicted by either model. The frequency dependency that is accounted for by the added mass coefficient is predicted by the models and arises (in the models) primarily because of the reduction in degrees of freedom from the initial 12 degrees (four pads times three degrees of freedom) to the two-rotor degrees of freedom. For the bearing and condition tested, pad and fluid inertia are secondary considerations out to running speed. The direct stiffness and damping coefficients increase with load, while increasing and decreasing with rotor speed, respectively. As expected, a small whirl frequency ratio (WFR) was found of about 0.15, and it decreases with increasing load and increases with increasing speed. The two model predictions for WFR are comparable and both underpredict the measured WFR values. Rotors supported by either conventional tilting-pad bearings or flexure-pivot tilting-pad (FPTP) bearings are customarily modeled by frequency-dependent stiffness and damping matrices, necessitating an iterative calculation for rotordynamic stability. For the bearing tested and the load conditions examined, the present results show that adding a constant mass matrix to the FPTP bearing model produces an accurate frequency-independent model that eliminates the need for iterative rotordynamic stability calculations. Different results may be obtained for conventional tilting-pad bearings (or this bearing at higher load conditions).*

[DOI: 10.1115/1.2179467]

Introduction

Figure 1 shows a flexible-pivot bearing (FPB). The pads are supported by a beam element "web." The web has sufficient radial stiffness to support the pad and the load of the shaft in the radial direction, and also allows the pads to tilt to achieve high stability (low cross-coupling stiffnesses). Armentrout and Paquette [1] introduced this design, which offers many of the rotordynamic advantages of tilting-pad journal bearings. FPB consists of pivoted pads machined from a solid blank through an electrical discharge machining (EDM) process. Performance data for a FPB in terms of stability were presented by Armentrout and Paquette [1] for a high-speed turbocompressor. Kepple et al. [2] reported that the first FPB set was installed in 1994. They compare some disadvantages of tilting-pad bearing (pivot wear, complexity in design, production, installation, tolerance stack ups, and maintenance) versus the FPBs that have stability characteristic similar to tilting-pad bearings. Zeidan [3] also discusses the rotordynamic characteristics of FPBs.

Stiffness, damping, and added-mass coefficients are very important in modeling fluid-structure interaction forces for rotordynamic analysis. These coefficients are typically expressed in the following linearized force-displacement bearing model:

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received August 30, 2005; final manuscript received September 6, 2005. Review conducted by K. C. Hall. Paper presented at the ASME Turbo Expo 2005: Land, Sea, and Air, Reno, NV, June 6–9, 2005, Paper No. GT2005-68343.

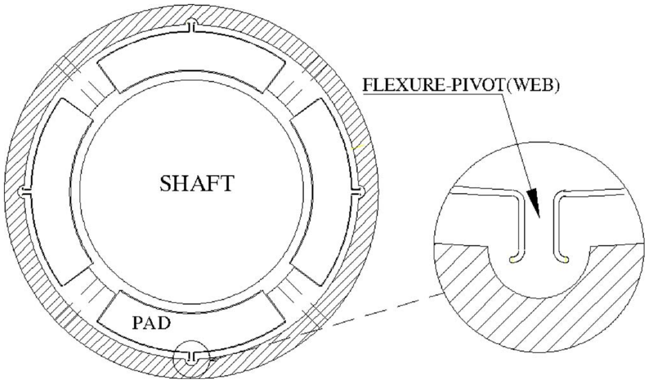


Fig. 1 Flexure-pivot tilting-pad bearings

$$-\begin{bmatrix} f_{bx} \\ f_{by} \end{bmatrix} = \begin{bmatrix} K_{xx} & K_{xy} \\ K_{yx} & K_{yy} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} + \begin{bmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{bmatrix} \begin{bmatrix} \Delta \dot{x} \\ \Delta \dot{y} \end{bmatrix} \tag{1}$$

In performing a stability analysis of high-speed machinery running on tilting-pad bearings, a persistent question is: Should the rotordynamic coefficients be calculated at the synchronous precession (running-speed) frequency or at the rotor's natural frequency? Lund [4] discusses how to calculate these coefficients at the running-speed (synchronous precession) frequency for a tilting-pad journal bearing, considering the pads' inertia and raised this question. His analysis was based on a Reynolds-equation model that does not predict added-mass coefficients for fixed-arc bearings. Reinhardt and Lund [5] show the significance of fluid inertia (added-mass coefficients) on a journal bearing operating in a laminar flow regime, producing the following model:

$$-\begin{bmatrix} f_{bx} \\ f_{by} \end{bmatrix} = \begin{bmatrix} K_{xx} & K_{xy} \\ K_{yx} & K_{yy} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} + \begin{bmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{bmatrix} \begin{bmatrix} \Delta \dot{x} \\ \Delta \dot{y} \end{bmatrix} + \begin{bmatrix} M_{xx} & M_{xy} \\ M_{yx} & M_{yy} \end{bmatrix} \begin{bmatrix} \Delta \ddot{x} \\ \Delta \ddot{y} \end{bmatrix} \tag{2}$$

San Andrés [6] presented an analysis of flexure-pivot tilting-pad hybrid bearings operating in the turbulent flow regime, using bulk-flow transport equations (mass-continuity, momentum, and energy equations). His predictions using liquid oxygen as a lubricant show both direct and cross-coupled added mass coefficients.

Parsell et al. [7] also obtained stiffness and damping coefficients for a tilting-pad bearing. They showed that the frequency of the excitation force is one of the important parameters affecting the dynamic characteristics of high-speed bearings. They concluded that the direct stiffness coefficients of a preloaded five-shoe tilting-pad bearing decrease slightly with an increase of excitation frequency, while the direct damping coefficients increase slightly. Barrett et al. [8] extended previous work by including the real parts of the system eigenvalues. Their results were presented for a five-shoe tilting-pad bearing (LBP (load-between-pads) configuration) with negligible pad inertia effects. They conclude that rotor stability calculations using synchronously reduced coefficients (ignoring real parts of the system eigenvalues) would overestimate the stability of the system for small preloads with high Sommerfeld numbers.

Ha and Yang [9] measured the effects of excitation frequency on the direct stiffness and direct damping coefficients of a five-shoe tilting-pad bearing (LOP (load-on-pad) configuration). The maximum rotational speed used in that test was 3600 rpm. They conclude that the variation of the excitation frequency has a negligible effect on both direct stiffness and damping coefficients.

For tilting-pad bearings, in general, and flexure-pivot pad bearings (FPB), in particular, the continuing rotordynamic issue is: What is the frequency-dependent nature of the rotordynamic coefficients? The present work answers this question by providing experimental test data for rotordynamic coefficients of a bearing for a range of running speeds, loads, and excitation frequencies. Test results are compared to predictions obtained from Reynolds-equation and bulk-flow models.

Test Rig Description

Figure 2 shows the main test section of the test rig described by Rodriguez and Childs [10] and used here to study the dynamic performance of the flexure-pivot tilting-pad bearing. The test rig consists of the following components:

1. a steel base to support the main test section and the air drive turbine
2. two pedestals to support the test rotor through two ball bearings that are lubricated through an oil-mist system
3. an air turbine that can deliver ~65 kW used to drive the rotor up to a maximum speed of 17,000 rpm through a high-speed flexible-disk coupling
4. a stator used to hold the flexure-pivot tilting-pad bearing and all the associated instrumentation (thermocouples, pressure probes, accelerometers, and proximity sensors)
5. six pitch stabilizers used to hold the stator axially; these pitch stabilizers are also used to angularly align the shaft with the rotor and to eliminate pitching motion.
6. a pneumatic loader to apply a static load to the stator in the positive *y* direction. The static loader cable is connected to the stator assembly through a yoke and spring system to assure that the load is applied exclusively in one direction. The rated maximum available load is ~22.2 kN. The applied load is measured with a load cell attached to the static loader cable.
7. two orthogonally mounted hydraulic shakers used mainly for applying dynamic load to the stator; their heads are attached to the stator middle section via elastic stingers. Stingers isolate the test structure from the dynamics of the shakers structure. The stator-shaker-stinger arrangement is shown in Fig. 3. The *x*-direction shaker can excite the stator with dynamic loads up to 4.45 kN in tension and compression, while the *y*-direction shaker can excite the stator with dynamic loads up to 4.45 kN in tension and 11.1 kN in compression. Both shakers can excite at frequencies up to 1000 Hz. The dynamic load applied to the stator is measured with load cells located between the stingers and the shaker frame.
8. oil supplying system. ISO VG32 turbine oil is delivered to the test section at the required specific temperature and flow rate through a controlled pneumatically driven valves and heat exchanger system.

Instrumentation

Six high-sensitivity proximity probes, located in the stator end caps on the exit sides of the FPB, record the relative motion of the stator with respect to the rotor for the *x* and *y* directions of excitation force. Two radial proximity probes were installed in the end cap at the non-drive-end (NDE) side. In addition, four probes were installed in the end cap at the drive-end (DE) side to measure the pitch and yaw of the stator. Two of the probes installed in the end cap at the DE side were used to provide feedback to the shakers' control system. Two piezoelectric accelerometers measure the stator absolute acceleration in the *x* and *y* directions.

Flexure-Pivot Bearing Geometry

Figure 4 shows the FPB-stator assembly, consisting of the bearing, stator, and two end caps. The FPB design parameters are given in Table 1.

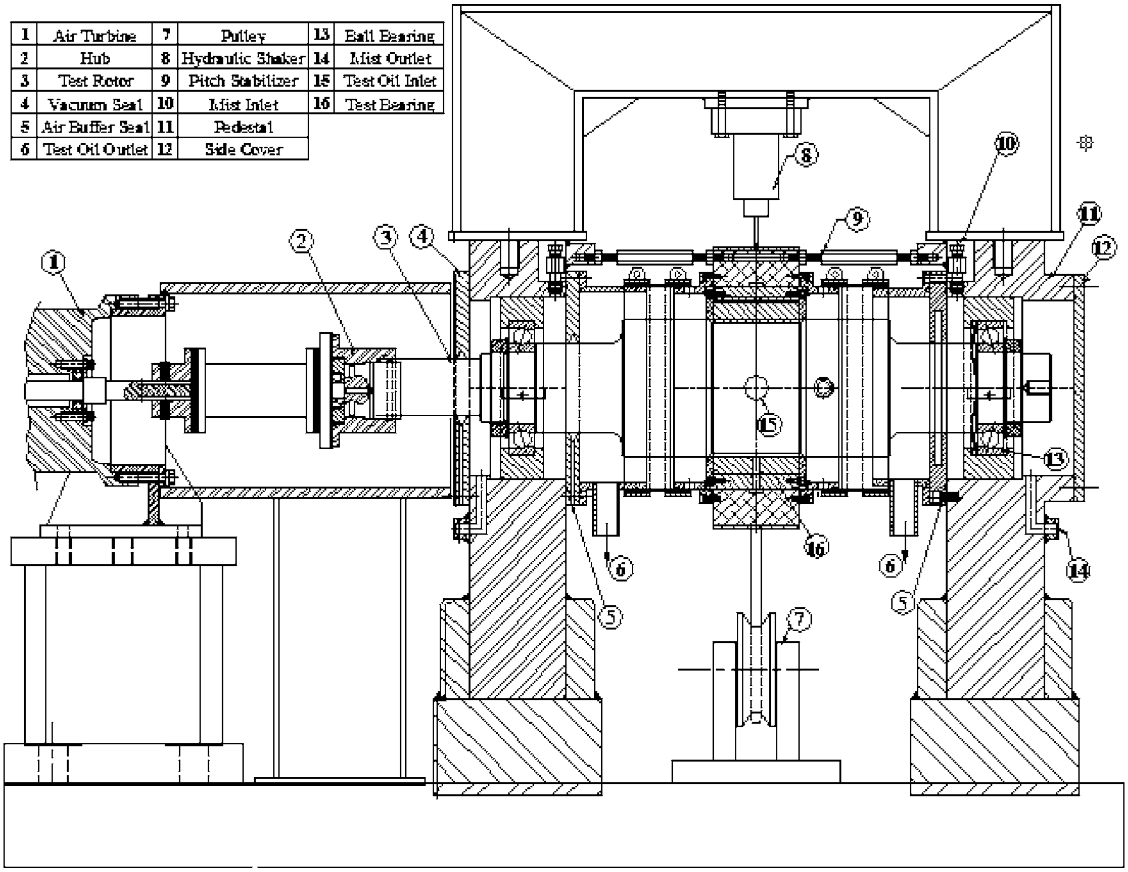


Fig. 2 Main test section of the test rig

Dynamic Parameters Identification Procedure

This section details the rotordynamic parameter identification procedure and has been adapted from Childs and Hale [11] and repeated here for clarification purposes. The stator assembly is excited by hydraulic shakers in two orthogonal planes (x,y). The equations of motion for the stator can be written as

$$M_s\begin{bmatrix} a_x \\ a_y \end{bmatrix} = \begin{bmatrix} f_x \\ f_y \end{bmatrix} - \begin{bmatrix} f_{bx} \\ f_{by} \end{bmatrix} \tag{3}$$

Substituting Eq. (2) into Eq. (3) and rearranging gives

$$\begin{bmatrix} f_x - M_s a_x \\ f_y - M_s a_y \end{bmatrix} = -\begin{bmatrix} K_{xx} & K_{xy} \\ K_{yx} & K_{yy} \end{bmatrix}\begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} - \begin{bmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{bmatrix}\begin{bmatrix} \Delta \dot{x} \\ \Delta \dot{y} \end{bmatrix} - \begin{bmatrix} M_{xx} & M_{xy} \\ M_{yx} & M_{yy} \end{bmatrix}\begin{bmatrix} \Delta \ddot{x} \\ \Delta \ddot{y} \end{bmatrix} \tag{4}$$

The rotordynamic coefficients are determined in the frequency domain via the fast Fourier transform $\mathcal{F}$, which yields

$$\begin{bmatrix} \boldsymbol{F}_x - M_s\boldsymbol{A}_x \\ \boldsymbol{F}_y - M_s\boldsymbol{A}_y \end{bmatrix} = -\begin{bmatrix} \boldsymbol{H}_{xx} & \boldsymbol{H}_{xy} \\ \boldsymbol{H}_{yx} & \boldsymbol{H}_{yy} \end{bmatrix}\begin{bmatrix} \boldsymbol{D}_x \\ \boldsymbol{D}_y \end{bmatrix} \tag{5}$$

where $\boldsymbol{F}_k=\mathcal{F}(f_k)$, $\boldsymbol{A}_k=\mathcal{F}(a_k)$, $\boldsymbol{D}_k=\mathcal{F}(\Delta k)$.

The elements of the frequency response function $\boldsymbol{H}$ in Eq. (5) are related to the coefficients defined in Eq. (4) by

$$\boldsymbol{H}_{ij} = K_{ij} - M_{ij}\Omega^2 + \boldsymbol{j}C_{ij}\Omega \tag{6}$$

Equation (5) provides two equations in the four unknowns $\boldsymbol{H}_{xx}$, $\boldsymbol{H}_{xy}$, $\boldsymbol{H}_{yx}$, and $\boldsymbol{H}_{yy}$. To provide four independent equations, alternate shakes about an eccentric rotor position were conducted on the stator in orthogonal directions (x and y) yielding four equations and four unknowns, given by

$$\begin{bmatrix} \boldsymbol{F}_{xx} - M_s\boldsymbol{A}_{xx} & \boldsymbol{F}_{xy} - M_s\boldsymbol{A}_{xy} \\ \boldsymbol{F}_{yx} - M_s\boldsymbol{A}_{yx} & \boldsymbol{F}_{yy} - M_s\boldsymbol{A}_{yy} \end{bmatrix} = -\begin{bmatrix} \boldsymbol{H}_{xx} & \boldsymbol{H}_{xy} \\ \boldsymbol{H}_{yx} & \boldsymbol{H}_{yy} \end{bmatrix}\begin{bmatrix} \boldsymbol{D}_{xx} & \boldsymbol{D}_{xy} \\ \boldsymbol{D}_{yx} & \boldsymbol{D}_{yy} \end{bmatrix} \tag{7}$$

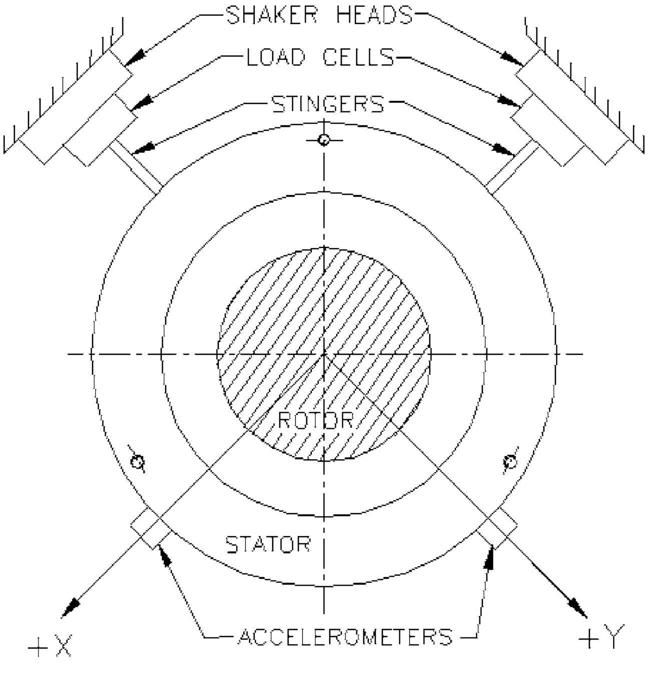


Fig. 3 Shaker-stinger configuration (NDE side)

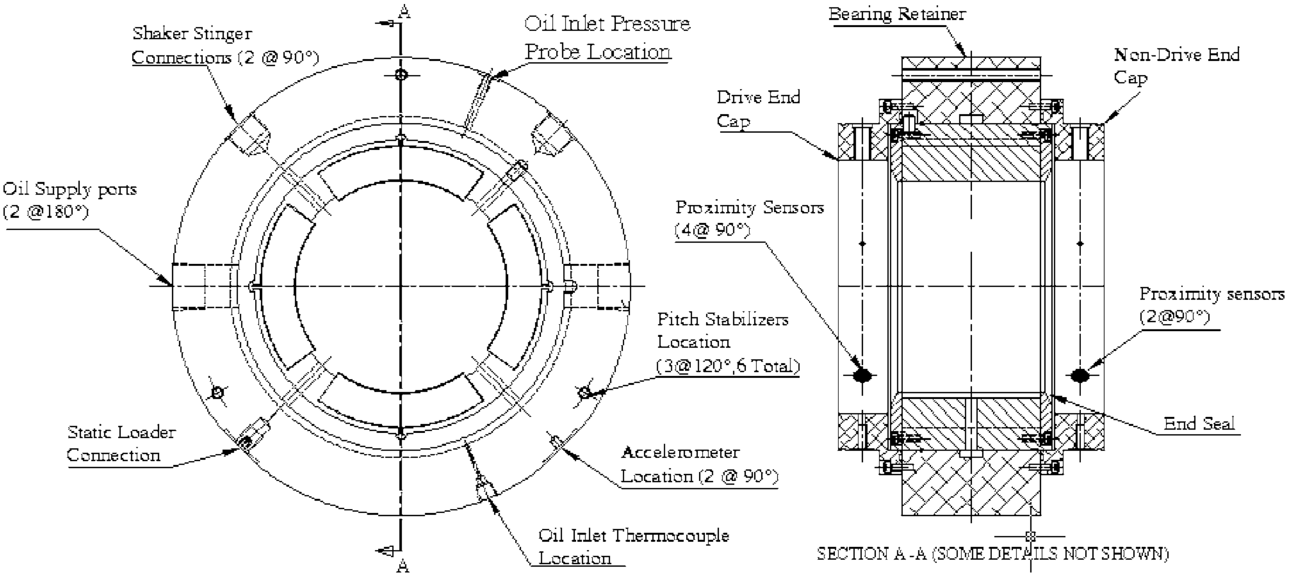


Fig. 4 Bearing-stator assembly

Bearing Center

The bearing center was found using "bump" tests. Diametral clearance was measured along two perpendicular axes passing through the center of the pads and found to be 330.2 μm and 431.8 μm, respectively. The difference suggests that the bearing was "crushed" at some stage of the assembly process.

Dynamic Tests Procedure

Dynamic tests were performed by shaking the stator in the x and y directions, independently, for every point in the test matrix shown in Table 2.The excitation of the shakers was a pseudo-random wave-form type with frequencies from 20 Hz to 320 Hz in 10 Hz increments. The magnitude of the excitation force was controlled such that the FPB will have a linear behavior during testing. In general, the motion of the stator has been kept to ~5–10% of the bearing radial clearance.

Table 1 Flexure-pivot bearing design parameters

Number of pads	4
Configuration	LBP
Pad arc angle	72°
Pivot offset	50%
Rotor diameter	116.8095±0.0051 mm
Pad axial length	76.2±0.0254 mm
Radial pad clearance (C_p)	0.254±0.0127 mm
Radial bearing clearance (C_b)	0.1905±0.0127 mm
Preload	0.25
Calculated pad rotational stiffness	1694.8 N m/rad
Calculated pad polar inertia	7.448×10^{-5} kg m^2
Calculated pad mass	1.226 kg
Web thickness	2.1251 mm
Web height	7.4379 mm

Table 2 Nominal test matrix for dynamic tests

	Bearing unit loading (kPa)					
rpm	0	172.4	344.7	517.1	689.5	1034.2
4000	*	*	*	*	*	N/A
6000	*	*	*	*	*	*
8000	*	*	*	*	*	*
10,000	*	*	*	*	*	*
12,000	*	*	*	*	*	*

One set of the dynamic stiffness coefficients is obtained as the average of 32 separate shake tests that are averaged in the frequency domain using a pseudo-random wave form. To estimate the variability of dynamic data, 10 consecutive tests (each including 32 separate shake tests) were conducted. During these tests, the operating conditions load, speed and temperature (37.8°C) were held approximately constant. The following sets of data were recorded at each of the test conditions shown in Table 2 in the time domain for a 3.2 s duration of shaking in each direction:

(a) excitation force component vector
(b) stator acceleration vector component
(c) stator position in the x and y directions at the NDE plane

The average of these 10 tests was used for calculating the dynamic stiffnesses. Their uncertainties are equal to two times the standard deviation. The dynamic tests aim to measure rotordynamic coefficients of the FPB. However, the procedure will also measure stiffness and damping arising from the structure of the test rig that holds or feeds the FPB, such as the hose connections, pitch stabilizers, etc. To account for these additional elements, "base-line" tests were conducted at zero rotor speed without oil in the bearing. The baseline dynamic stiffnesses have been subtracted from the average dynamic stiffnesses obtained for each test condition, and the rotordynamic coefficients are obtained from the corrected dynamic stiffness result.

Software based on San Andrés [12] was used to predict the static and dynamic performance of the FPB. This model is based on the bulk-flow governing equations (mass conservation, axial and circumferential momentum, and energy equations) for laminar and turbulent flow in the thin fluid film lands of the bearing. Fluid inertia effects (temporal and convective terms) can be accounted for using this software. The highest Reynolds number calculated in these experiments was <350 at 12,000 rpm, which indicates operation in the laminar flow regime. A Reynolds equation model developed by San Andres was also used. Pad inertia and pad support flexibility is accounted for in all calculations for both models. An isothermal analysis is used. The codes cannot account for dif-

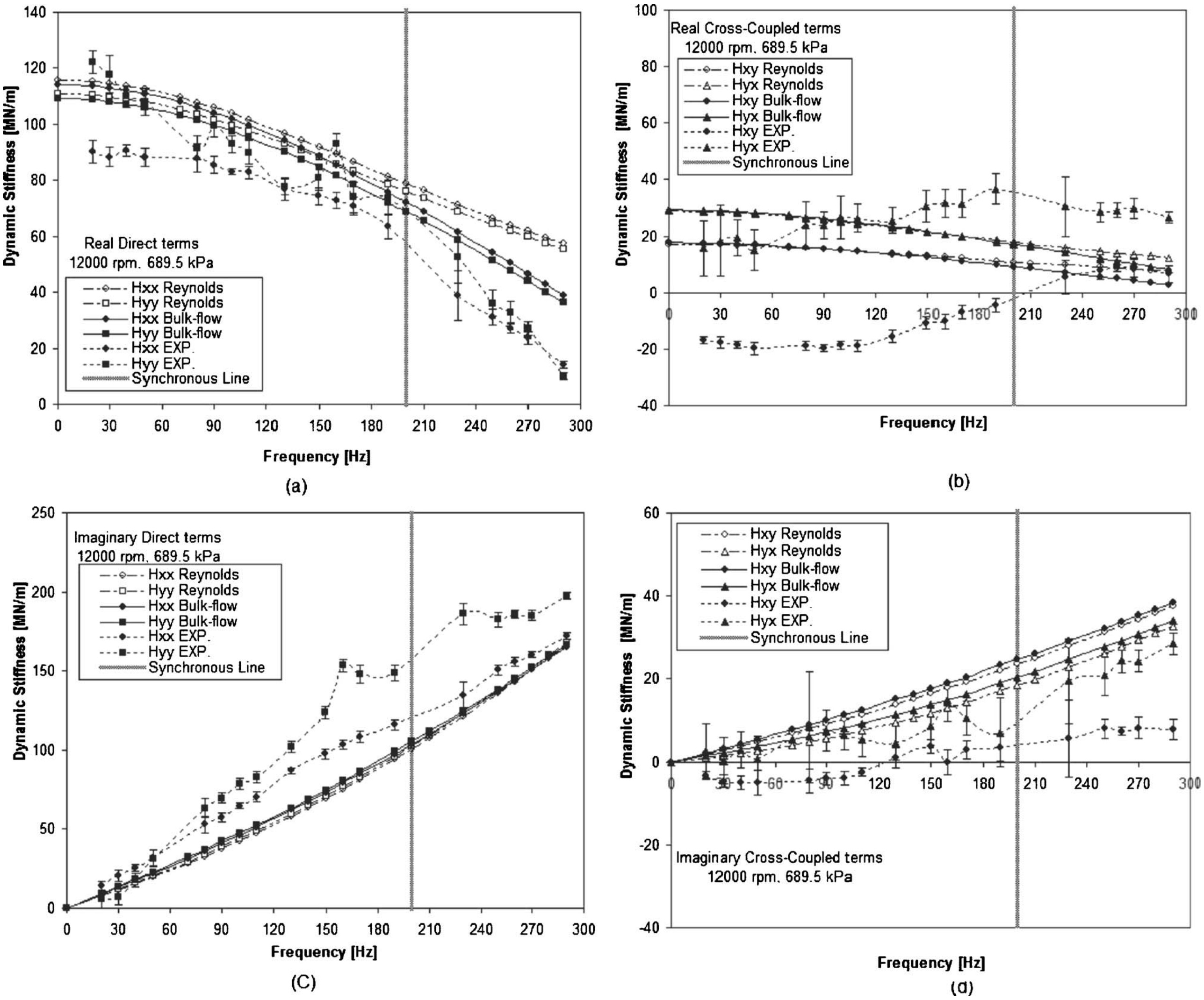


Fig. 5 Experimental and theoretical dynamic stiffnesses versus the excitation frequency at 12,000 rpm and 689.5 kPa bearing unit loading

ferent radial clearances in orthogonal directions so a constant radial clearance of 165.1 μm (6.5 mils) has been used for all calculations.

Identifying Rotordynamic Coefficients

After subtracting the baseline results, the average dynamic stiffnesses can be written in a matrix form given by

$$\begin{bmatrix} \boldsymbol{H}_{xx} & \boldsymbol{H}_{xy} \\ \boldsymbol{H}_{yx} & \boldsymbol{H}_{yy} \end{bmatrix} = \begin{bmatrix} \mathrm{Re}(\boldsymbol{H}_{xx}) & \mathrm{Re}(\boldsymbol{H}_{xy}) \\ \mathrm{Re}(\boldsymbol{H}_{yx}) & \mathrm{Re}(\boldsymbol{H}_{yy}) \end{bmatrix} + j \cdot \begin{bmatrix} \mathrm{Im}(\boldsymbol{H}_{xx}) & \mathrm{Im}(\boldsymbol{H}_{xy}) \\ \mathrm{Im}(\boldsymbol{H}_{yx}) & \mathrm{Im}(\boldsymbol{H}_{yy}) \end{bmatrix} \tag{8}$$

The rotordynamic coefficients can be calculated from the frequency response functions $\boldsymbol{H}$ by performing a least-squares parameter estimation on each element of the real and the imaginary parts of Eq. (8). The real and the imaginary parts take the forms $K_{ij}-M_{ij}\Omega^2$ and $C_{ij}\Omega$, respectively, where Ω is the excitation frequency. Therefore, the stiffness and added-mass coefficients will be estimated from the real part of Eq. (8); whereas the damping is estimated from the slope of the imaginary part of Eq. (8). The uncertainty of the rotordynamic coefficients can be found utilizing the confidence interval for that particular coefficient; a 95% confidence interval was used, which mainly depends on the curve fits of the average value of the dynamic stiffnesses in Eq. (8).

The predicted dynamic stiffnesses were treated the same way as the experimental dynamic stiffnesses in obtaining the rotordynamic coefficients and their uncertainties.

Dynamic Stiffness Results

Figure 5 presents the experimental and theoretical (bulk-flow and Reynolds-equation) dynamic stiffness coefficients at 12,000 rpm and 689.5 kPa bearing unit loading. These results are representative of all dynamic-stiffness data. Results are shown for $\boldsymbol{H}_{xx}$, $\boldsymbol{H}_{yy}$, $\boldsymbol{H}_{xy}$, and $\boldsymbol{H}_{yx}$ for a range of experimental frequencies (20–320 Hz). The running speed is indicated by the vertical line at 200 Hz.

Starting with Re($\boldsymbol{H}_{xx}$) and Re($\boldsymbol{H}_{yy}$) in Fig. 5(*a*), remember that the zero-frequency intercept defines the direct stiffness. The two model predictions and the measurements show a very substantial drop in stiffness with increasing excitation frequency Ω. This quadratic frequency dependency can be modeled as an added-mass coefficient. Both models give comparable results below the running speed, but the bulk-flow predictions are modestly closer to test results than the Reynolds solution, especially at high frequen-

Table 3 Coefficient of determination for experimental dynamic stiffnesses shown in Fig 5

	Re($\boldsymbol{H}_{xx}$)	Re($\boldsymbol{H}_{xy}$)	Re($\boldsymbol{H}_{yx}$)	Re($\boldsymbol{H}_{yy}$)	Im($\boldsymbol{H}_{xx}$)	Im($\boldsymbol{H}_{xy}$)	Im($\boldsymbol{H}_{yx}$)	Im($\boldsymbol{H}_{yy}$)
R^2	0.9852	0.9269	0.3136	0.9360	0.9962	0.9033	0.8533	0.9578

cies. These differences will result in smaller direct added-mass coefficients for the Reynolds equation model. Re($\boldsymbol{H}_{yy}$) is greater than Re($\boldsymbol{H}_{xx}$) because +y is the load direction, and the bearing has a tighter clearance in that direction (bearing crush). A conventional stiffness and damping matrix model for the FPB will require a quadratic dependency of the direct stiffness coefficient to account for the measured and predicted results. Simply adding a direct added mass coefficient to the model, as shown in Eq. (3), eliminates the frequency dependency.

Moving to Re($\boldsymbol{H}_{xy}$) and Re($\boldsymbol{H}_{yx}$) in Fig. 5(b), results and predictions define frequency-dependent cross-coupled stiffness coefficients. There are slight differences in predictions from the two models at high (supersynchronous) frequencies, but the results basically coincide below the running speed. Note, in particular, that the measured values for Re($\boldsymbol{H}_{xy}$) and Re($\boldsymbol{H}_{yx}$) have difference signs in the subsynchronous frequency range versus other predicted terms that have the same signs, implying that the measured cross-coupled stiffness coefficients are more destabilizing than predicted. The erratic fluctuations with excitation frequency are not predicted by either model.

Continuing with Im($\boldsymbol{H}_{xx}$) and Im($\boldsymbol{H}_{yy}$) in Fig. 5(c), the slope of this function defines the direct damping coefficient. As predicted, measured data generally increases linearly with increasing frequency. Predictions from the two models coincide, and the measured slope is consistently larger than predicted. As the next section concerning rotordynamic coefficients shows, this outcome varies with load and speed. For some ranges of these parameters, damping is underpredicted.

Figure 5(d) illustrates Im($\boldsymbol{H}_{xy}$) and Im($\boldsymbol{H}_{yx}$). Predictions from the two models basically coincide and predict a linear increase with increasing frequency. The slopes of the functions define the cross-coupled stiffness coefficients of Eq. (4). Measured results, over the full frequency range, are nominally linear: However, a good deal of unpredicted (and unexplained) frequency dependency is present in the measured results. Cross-coupled damping coefficients that are approximately equal and have the same sign provide positive damping. A comparison between the magnitudes in Figs. 5(c) and 5(d) shows that the cross-coupled damping is approximately one-fifth of the direct damping. Predicted and extracted rotordynamic coefficients will be presented and discussed in the next section.

For some test condition, the parabolic $K_{ij}-M_{ij}\Omega^2$ and the linear $C_{ij}\Omega$ models used to fit the real and imaginary cross-coupled dynamic stiffnesses worked poorly, as can be seen from the coefficient of determination R^2 values in Table 3. Specifically for the Re($\boldsymbol{H}_{yx}$) in this example R^2 values equal to 1 represents a perfect fit, and lower values indicate a poorer fit. In general, the direct dynamic stiffnesses always have a good fit unlike the cross-coupled terms.

Rotordynamic Coefficients Result

Stiffness Coefficients. Figure 6 shows the experimental and theoretical (bulk-flow model) stiffness coefficients versus rotor speed for different bearing unit loading. The direct stiffness increases linearly with increasing rotor speed at low bearing unit loading as seen in Fig. 6(a). As seen in Fig. 6(e), increasing the bearing unit loading causes the direct stiffness coefficients to become less sensitive to increases in rotor speed. When increasing the load further to the maximum loading (9.2 kN), Figs. 6(e) and 6(f), the experimental direct stiffness coefficients increase with increasing rotor speed up to 10,000 rpm and then decrease for speeds above 10,000 rpm.

Direct stiffness coefficients are well predicted at low loads, but at the highest loads are underpredicted. Also, at higher loads, measured direct stiffnesses are more dependent on speed than predicted. The cross-coupled stiffness coefficient magnitudes increase uniformly with increasing rotor speed. The maximum and minimum increase of cross-coupling stiffness coefficient K_{xy} is seen, respectively, at the lowest load in Fig. 6(a) and at the maximum load in Fig. 6(f), while for K_{yx} the opposite situation holds. Contrary to predictions, K_{yx} and K_{xy} have different signs for all operating conditions. The agreement between experiment and predictions improves with increasing rotor speed and decreasing applied static load.

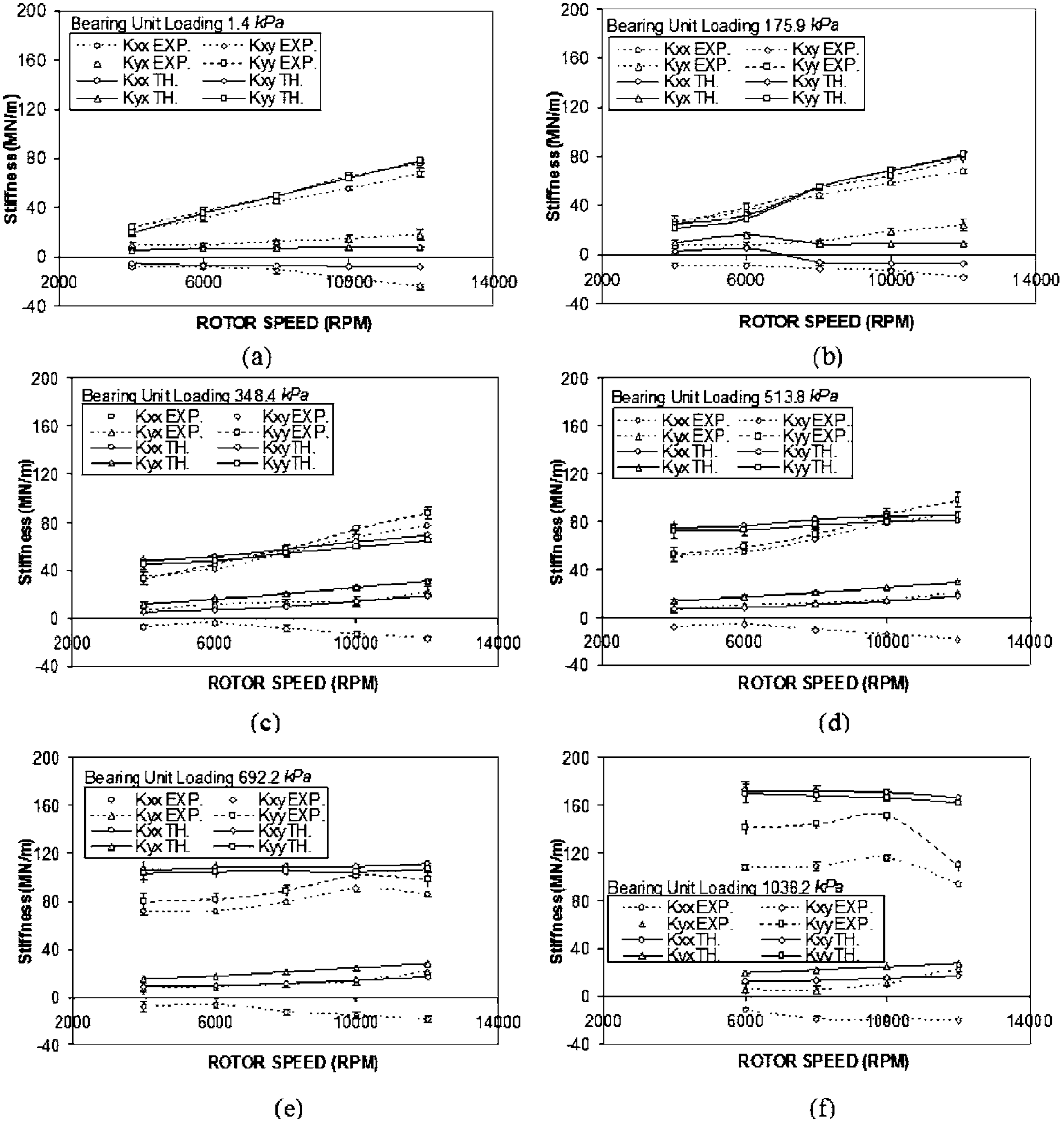
Damping Coefficients. Figure 7 illustrates measured and predicted direct and cross-coupled damping coefficients for a range of applied loads. Figure 7 shows that the direct damping coefficient in the direction of load C_{yy} and the cross-coupled damping coefficient C_{yx} are always higher than C_{xx} and C_{xy}, respectively, because the clearance in the y direction is smaller than in the x direction (bearing crush discussed earlier). Direct damping coefficients C_{yy} and C_{xx} increase with increasing static load. For increasing loads, they drop with increasing speed. Direct damping coefficients are reasonably well predicted, with better accuracy at low loads than high loads.

Added-Mass Coefficients. Figure 8 illustrates the added-mass coefficients for a range of running speeds and applied bearing unit loading. At low unit loads, M_{yy} and M_{xx} are approximately equal (32 kg) and independent of speed. With increasing unit loads, M_{xx} is smaller than M_{yy} at low speeds but approaches M_{yy} at high speeds. The cross-coupled added mass coefficients M_{yx} and M_{xy} are relatively small, tend to both be negative, and are insensitive to changes in speed or unit loading. The direct added-mass terms are reasonably well predicted at low and medium loads except for the major differences in magnitude for M_{yy} and M_{xx}. Measured cross-coupled mass coefficients have about the same magnitudes as predictions and, generally, have the same sign (as predicted), but their signs are incorrectly predicted.

Negative values for cross-coupled added mass terms come from the approach used to fit the dynamic stiffnesses to the $K_{ij}-M_{ij}\Omega^2$ model. The frequency dependency of measured cross-coupled dynamic stiffnesses does not follow this model. But, the cross-coupled terms were very small compared to the direct coefficients, and to maintain a consistency in presenting the data, these coefficients were not modified, even though the coefficient of determination (R^2) is very small for the curve fit.

Whirl Frequency Ratio (WFR)

WFR is the ratio between the precession frequency of the rotor and running speed at the onset speed of instability (ω_s). Lund [13] shows how to calculate the WFR for a rigid or flexible rotor supported by journal bearings that are modeled by stiffness and damping coefficients. To account for the effects of an added mass matrix, San Andres' analysis [14] was used. The WFR is a real positive number or purely imaginary number (indication of infinite stability), and in that case the WFR, will be a negative number. For most fixed-arc bearings, the WFR is ~0.5. As mentioned earlier, FPB characteristics lie between fixed-geometry bearings and tilting-pad bearings, depending on the rotational stiffness of



(a) (b) (c) (d) (e) (f)

Fig. 6 Stiffness coefficients versus rotor speeds for different bearing unit loading: (*a*) 1.4 kPa, (*b*) 175.9 kPa, (*c*) 348.4 kPa, (*d*) 513.8 kPa, (*e*) 692.2 kPa, and (*f*) 1038.2 kPa

the web. If the web rotational stiffness is high, the FPB will act like a fixed-arc bearing journal bearing. If the rotational stiffness is very small, the FPB will act like a tilting-pad bearing. Zeidan and Paquette [15] discuss in detail the effects of the flexure-pivot rotational stiffness on stability. The central advantage of the WFR in evaluating bearing stability characteristics is that the effects of all elements are considered.

Figure 9 shows WFR versus running speed for a range of applied loads. Increasing the load generally decreases the WFR. For the highest load, WFR is negative (infinite stability) at all speeds below 12,000 rpm. At all loads, the WFR approaches 0.15 at 12,000 rpm.

Figure 10 provides a WFR comparison between measurements and predictions (bulk-flow and Reynolds) models for light (1.4 kPa) and heavy (692.2 kPa) unit loads versus rotor speed. The models underpredict WFR (over estimate bearing stability) for both loading conditions. Bulk-flow predictions are modestly closer to measurement results at low loading. At higher loading, results from the two models are quite close. In addition, both models incorrectly predict the general trend of WFR with increasing rotor speed.

Summary and Conclusion

The following question was posed at the beginning of this paper: What is the frequency-dependent nature of the rotordynamic coefficients? To answer the question, tests were conducted for a FPB bearing over a range of unit loads (1.4–1038 kPa) and running speeds (4000–12,000 rpm). Measured excitation forces were

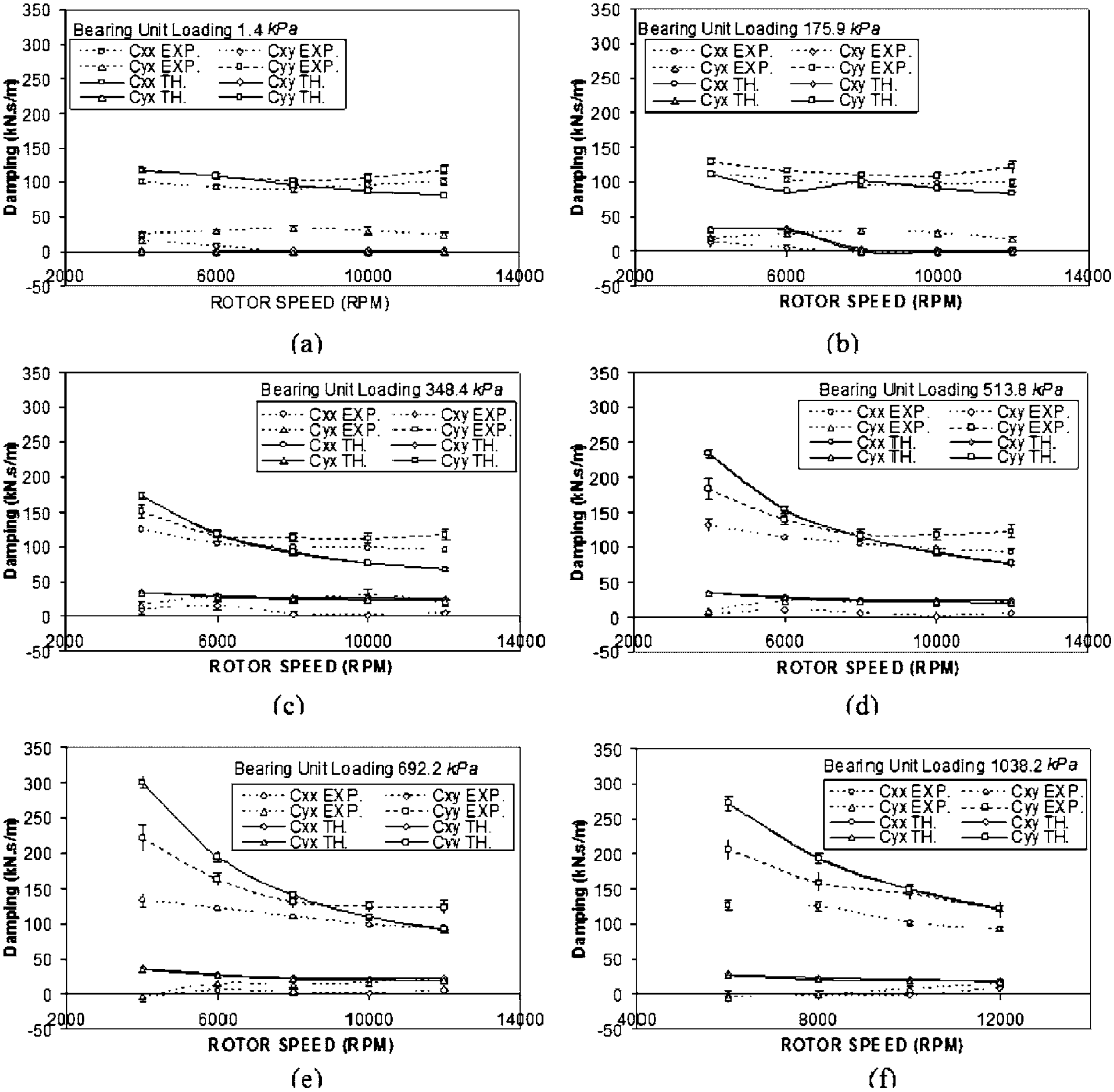


Fig. 7 Damping coefficients versus rotor speeds for different bearing unit loading: (*a*) 1.4 kPa, (*b*) 175.9 kPa, (*c*) 348.4 kPa, (*d*) 513.8 kPa, (*e*) 692.2 kPa, and (*f*) 1038.2 kPa

applied in two orthogonal directions at frequencies from 20 Hz to 320 Hz in two orthogonal directions to obtain direct and cross-coupled dynamic stiffness coefficients, from which rotordynamic coefficients were extracted.

The results obtained show that the answer to the frequency-dependent question depends entirely on the model used in defining the bearings rotordynamic coefficients. If a conventional stiffness and damping matrix model ([C] and [K] model) is used, the direct stiffness coefficients will be strongly frequency dependent, with a magnitude that drops dramatically with increasing excitation frequency. However, if a mass matrix model is added ([M]-[C]-[K] model), the model is frequency independent. This result holds for predictions from either a Reynolds-equation or a bulk-flow analysis. The bulk-flow model is modestly better at higher frequencies. However, out to running speed, the predictions from the two models basically coincide. The [M]-[C]-[K] model results may or may not apply to conventional tilting-pad bearings or even to the present bearing at higher load conditions. However, very similar results were obtained for this bearing in load and pad configuration [10]. Tests were conducted in 2005 for this flexure pivot pad bearing in LBP configuration at higher loads and a conventional tilting-pad bearing in LOP and LBP configurations.

The following points summarize and conclude the additional results that were obtained:

- Each element of the dynamic stiffness matrix $\boldsymbol{H}$ was fitted to $K_{ij}-M_{ij}\Omega^2+\boldsymbol{j}C_{ij}\Omega$ to obtain rotordynamic coefficients. A goodness of fit (coefficient of determination R^2) was calculated for each curve fit. The goodness of fit was found to be excellent for the direct coefficients and in some cases poor for the cross-coupled coefficients. Therefore the cross-coupled dynamic stiffness terms show a frequency dependency that cannot be fitted to $K_{ij}-M_{ij}\Omega^2+\boldsymbol{j}C_{ij}\Omega$.

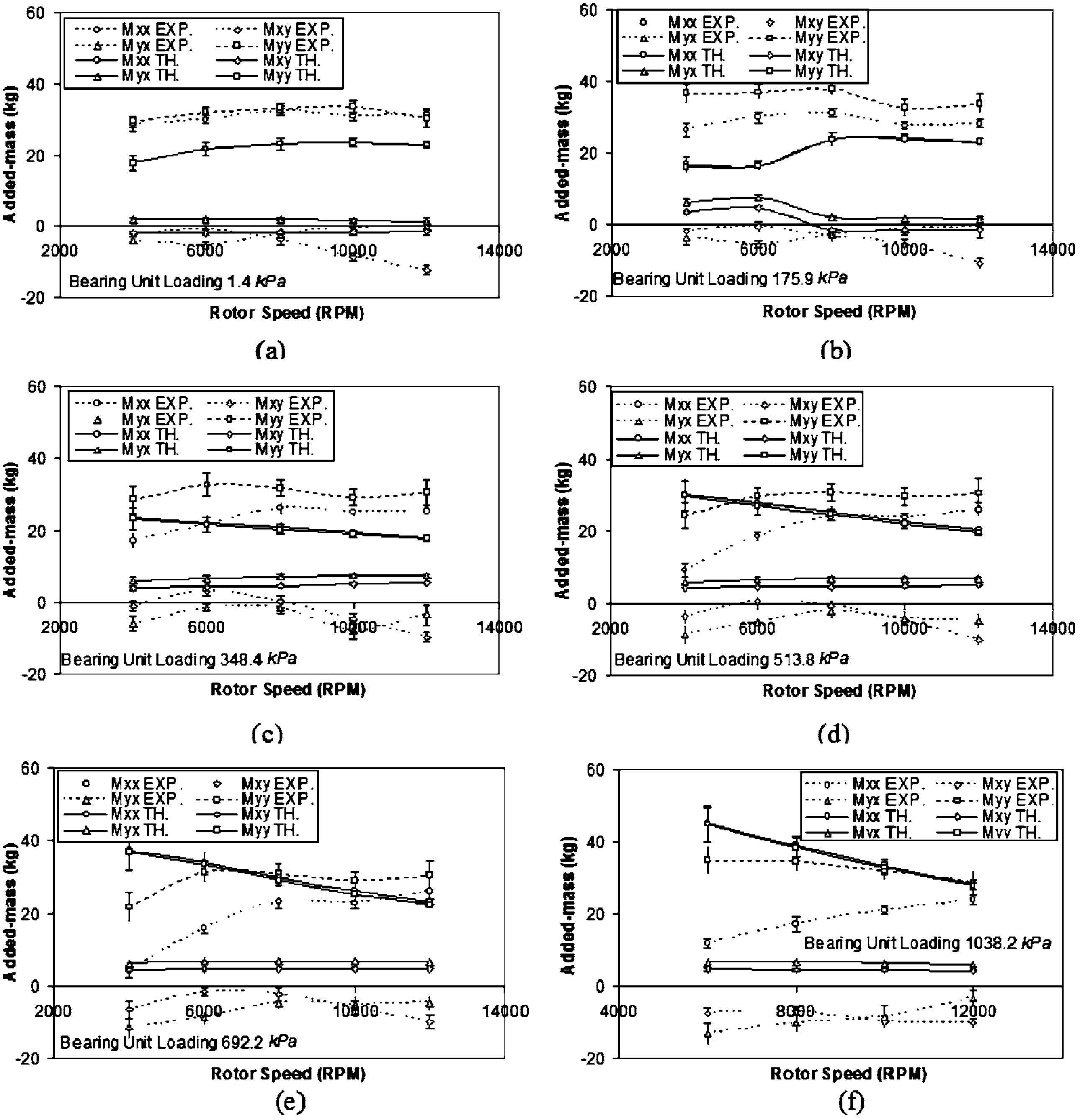


Fig. 8 Added-mass coefficients versus rotor speed for different bearing unit loading: (*a*) 1.4 kPa, (*b*) 175.9 kPa, (*c*) 348.4 kPa, (*d*) 513.8 kPa, (*e*) 692.2 kPa, and (*f*) 1038.2 kPa

- The direct stiffness coefficients increase with load and increase with speed at low loads. The direct damping coefficients increase with increasing load and decrease with increasing speed, with the maximum decrease at high loads.
- The direct added mass coefficients are almost equal ($M_{xx} \cong M_{yy}$) and on the order of (32 kg). They are almost constant with load specifically at high speeds, but increase with speed.
- The WFR of the flexure-pivot bearing (FPB) was calculated and found to be ~0.15. The WFR decreases with increasing load and increases with increasing speed with minimum value around 6000 rpm. The added mass coefficients have very little influence on the WFR, with ~1% increase at the highest speed and lowest load. The models predict infinite stability at low loads, while the measurements show a whirl frequency ratio on the order of 0.15. This would translate into a prediction of an onset speed of instability of 6.7 times the first critical speed versus infinity for the models.

Nomenclature

A_x, A_y = Fourier transformation of a_x and a_y, respectively
a_x, a_y = measured stator acceleration in the x and y directions, respectively (m/s^2)
C_b = radial bearing clearance (m)

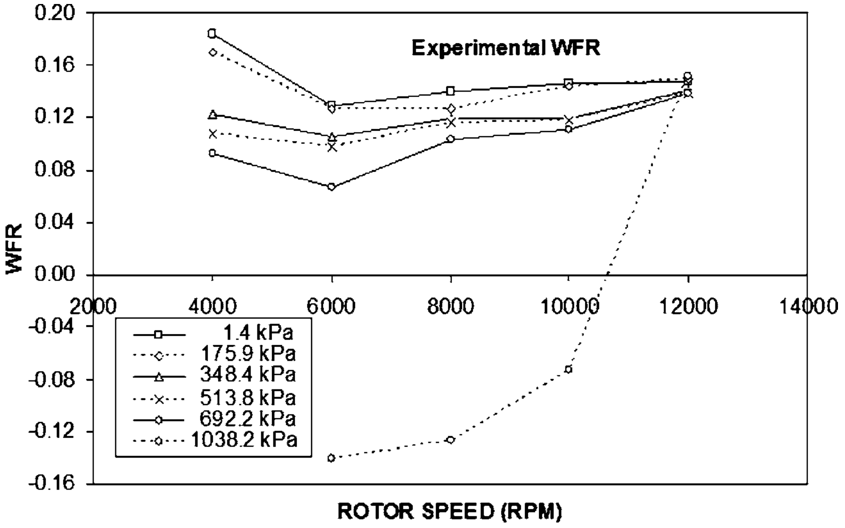


Fig. 9 Experimental WFR versus for rotor speed for different bearing unit loading

C_p = radial pad clearance (m)
C_{ij} = damping coefficient (kN s/m)
D = inside bearing diameter (m)
$\boldsymbol{D}_x, \boldsymbol{D}_y$ = Fourier transformation of Δx and Δy, respectively
$\boldsymbol{F}_x, \boldsymbol{F}_y$ = Fourier transformation of f_x and f_y, respectively
f_{bx}, f_{by} = bearing reaction force in the x and y directions, respectively (N)
f_x, f_y = measured excitation force in the x and y directions, respectively (N)
$\boldsymbol{H}_{ij}$ = $K_{ij} - M_{ij}\Omega^2 + jC_{ij}\Omega$, Average dynamic stiffness vector for the ten tests (MN/m)
i, j = subscripts representing x and y
$\boldsymbol{j}$ = $\sqrt{-1}$
K_{ij} = stiffness coefficient (MN/m)
L = pad length (m)
M_{ij} = added mass coefficient (kg)
M_s = stator mass (kg)
p = bearing unit loading $= W/(LD)$ (kPa)
R_b = bearing radius (m)
R_p = pad radius (m)
R_s = shaft radius (m)
W = applied static load in the positive y direction (N)
$\Delta x, \Delta y$ = measured relative displacement between the rotor and the bearing in the x and y directions, respectively (m)
x, y = displacement direction
ω = rotor speed (rpm)
Ω = excitation frequency (Hz)
ω_s = onest speed of instability (rpm)

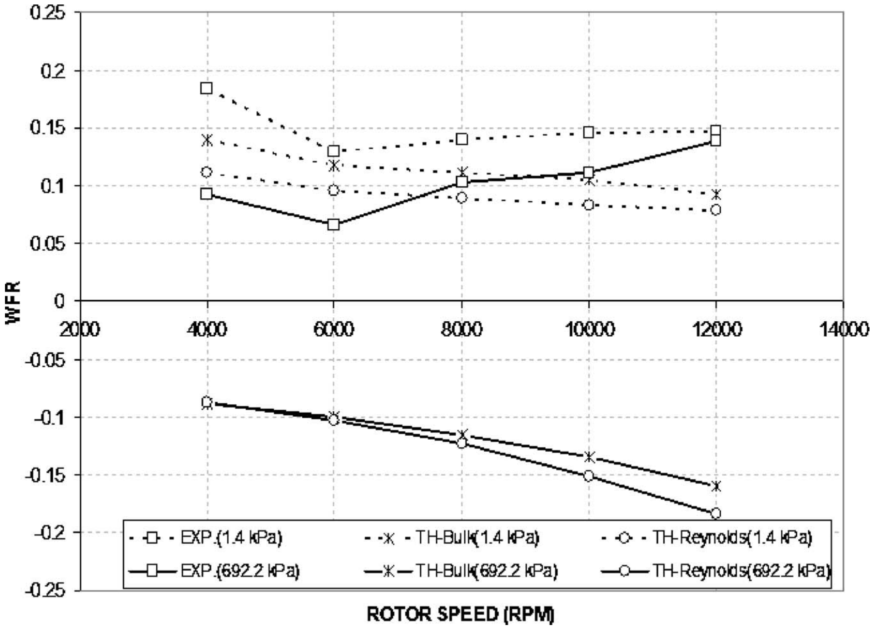


Fig. 10 Experimental and theoretical (bulk-flow and Reynolds models) WFR versus rotor speed for different bearing unit loading, with fluid inertial effects

References

[1] Armentrout, R. D., and Paquette, D. J., 1993, "Rotordynamic Characteristics of Flexure-Pivot Tilting-Pad Journal Bearings," Tribol. Trans., **36**, pp. 443–451.

[2] Kepple, W. E., Read, D. W., Zeidan, F. Y., Paraskevakos, C., and Dawson, M. P., 1998, "Experience in the Use of Flexure Pivot Tilt Pad Bearings in Boiler Feed Water Pumps," *Proceedings of the 15th International Pump Users Symposium*, Houston, Turbomachinery Laboratory, College Station, TX, pp. 77–84.

[3] Zeidan, F. Y., 1992, "Developments in Fluid Film Bearing Technology," Turbomach. Int., **9**, pp. 24–31.

[4] Lund, J., 1964, "Spring and Damping Coefficients for the Tilting Pad Journal Bearing," ASLE Trans., **7**, pp. 342–352.

[5] Reinhardt, E., and Lund, J., 1975, "The Influence of Fluid Inertia on the Dynamic Properties of Journal Bearings," ASME J. Lubr. Technol., **97**, pp. 159–167.

[6] San Andrés, L. A., 1996, "Turbulent Flow, Flexure-Pivot Hybrid Bearings for Cryogenic Applications," Trans. ASME, **118**, pp. 190–200.

[7] Parsell, J. K., Allaire, P. E., and Barrett, L. E., 1982, "Frequency Effects in Tilting-Pad Journal Bearing Dynamic Coefficients," ASLE Trans., **26**, pp. 222–227.

[8] Barret, L., Allaire, P., and Wilson, B., 1987, "The Eigenvalue Dependence of Reduced Tilting Pad Bearing Stiffness and Damping Coefficients," Tribol. Trans., **31**, pp. 411–419.

[9] Ha, H. C., and Yang, S. H., 1999, "Excitation Frequency Effects on the Stiffness and Damping Coefficients of a Five-Pad Tilting Pad Journal Bearing," ASME J. Tribol., **121**, pp. 517–522.

[10] Rodriguez, L., and Childs, D. W., 2004, "Experimental Rotordynamic Coefficient Results for a Load-on-Pad Flexible-Pivot Tilting-Pad Bearing With Comparisons to Predictions From Bulk-Flow and Reynolds Equation Models," ASME/STLE Joint Tribology Conference, ASME/STLE Paper No. TRIB2004-64042, Long Beach, CA, October.

[11] Childs, D. W., and Hale, K., 1994, "A Test Apparatus and Facility to Identify the Rotordynamic Coefficients of High-Speed Hydrostatic Bearings," ASME J. Tribol., **116**, pp. 337–344.

[12] San Andrés, L. A., 1995, "Bulk-Flow Analysis of Flexure and Tilting Pad Fluid Film Bearings," Turbomachinery Laboratory, Texas A&M University, College Station, Report No. TRC-B&C-3-95.

[13] Lund, J., 1965, "The Stability of an Elastic Rotor in Journal Bearings with Flexible Damped Supports," ASME J. Appl. Mech., **87**, pp. 911–920.

[14] San Andrés, L. A., 1991, "Effect of Eccentricity on the Force Response of a Hybrid Bearing," STLE Tribol. Trans., **34**(4), pp. 537–544.

[15] Zeidan, F. Y., and Paquette, D. J., 1994, "Application of High Speed and High Performance Fluid Film Bearings in Rotating Machinery," *Proceedings of the 23rd Turbomachinery Symposium*, Dallas, Turbomachinery Laboratory, College Station, TX, pp. 209–234.

A Comparative Study of Different Methods of Using Animal Fat as a Fuel in a Compression Ignition Engine

M. Senthil Kumar

A. Kerihuel

J. Bellettre[1]
e-mail: Jerome.bellettre@emn.fr

M. Tazerout

Département Systèmes Energétiques et
Environnement,
Ecole des Mines de Nantes,
4 rue Alfred Kastler,
BP 10722,
44307 Nantes,
Cedex 03, France

This work explores a comparative study of different methods of using animal fat as a fuel in a compression ignition engine. A single-cylinder air-cooled, direct-injection diesel engine is used to test the fuels at 100% and 60% of the maximum engine power output conditions. Initially, animal fat is tested as fuel at normal temperature. Then, it is preheated to 70°C *and used as fuel. Finally, animal fat is converted into methanol and ethanol emulsions using water and tested as fuel. A drop in cylinder peak pressure, longer ignition delay, and a lower premixed combustion rate are observed with neat animal fat as compared to neat diesel. With fat preheating and emulsions, there is an improvement in cylinder peak pressure and maximum rate of pressure rise. Ignition delay becomes longer with both the emulsions as compared to neat fats. However, preheating shows shorter ignition delay. Improvement in heat release rates is achieved with all the methods as compared to neat fats. At normal temperature, neat animal fat results in higher specific energy consumption and exhaust gas temperature as compared to neat diesel at both power outputs. Preheating and emulsions of animal fat show improvement in performance as compared to neat fat. Smoke is lower with neat fat as compared to neat diesel. It reduces further with all the methods. At peak power output, the smoke level is found as* $0.89\ \mathrm{m}^{-1}$ *with methanol,* $0.28\ \mathrm{m}^{-1}$ *with ethanol emulsions, and* $1.7\ \mathrm{m}^{-1}$ *with fat preheating, whereas it is* $3.7\ \mathrm{m}^{-1}$ *with neat fat and* $6.3\ \mathrm{m}^{-1}$ *with neat diesel. Methanol and ethanol emulsions significantly reduce NO emissions due to the vaporization of water and alcohols. However, NO increases with fat preheating due to high in-cylinder temperature. Higher unburned hydrocarbon and carbon monoxide emissions are found with neat fat as compared to neat diesel at both power outputs. However, these emissions are considerably reduced with all the methods. It is finally concluded that adopting emulsification with the animal fat can lead to a reduction in emissions and an improvement in combustion characteristics of a diesel engine.* [DOI: 10.1115/1.2180278]

Introduction

The main problems associated with the use of animal fats and vegetable oils as fuel in diesel engines are their high viscosity and poor volatility. A number of methods have been tried in the past to use vegetable oils and animal fats efficiently in diesel engines. Some of them are transesterification of vegetable oils, blending the oils with diesel and alcohols, fuel preheating, dual fueling with gaseous and liquid fuels, use of additives, etc. [1–5].

Transesterification shows significant reduction in viscosity, enhancement in cetane number, and other physical properties [5]. But, transesterification is a complex and cumbersome process. Dual fueling is a well-established technique to use different types of fuels in diesel engines [6]. A dual fuel engine results in good thermal efficiency and low smoke emissions, particularly at high power outputs [7]. However, dual fuel operation needs modification in the engine design. In addition, dual fuel operation with alcohol induction results in higher hydrocarbon and carbon monoxide emissions [7]. Though blending of oils with alcohols is a simple process, significant improvement in performance and reduction emissions are not reported in the literature. In addition, more quantities of alcohol addition to the oils leads to phase separation [8].

The fuel preheating technique offers the advantage of easy conversion of the normal diesel engine to work on highly viscous fuels. It needs no modification in the engine. Past investigations showed that preheated animal fat and vegetable oils in diesel engines resulted in improved brake thermal efficiency and reduced smoke and particulate emissions [4,9,10]. Since animal fats have very high viscosity, the preheating technique can offer significant reduction in viscosity with improved performance and reduced emissions in a diesel engine fueled with animal fats.

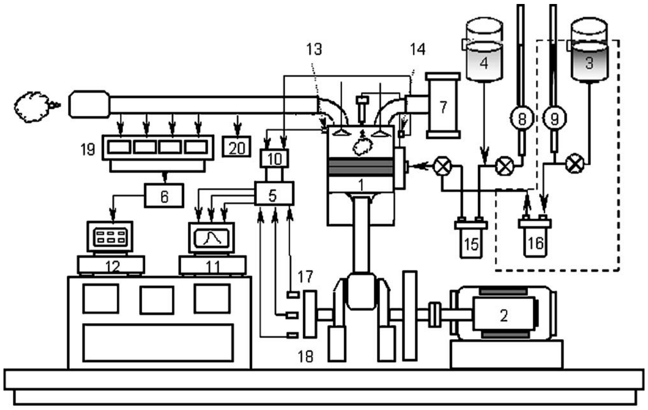
Emulsions also find attraction to use as fuel in diesel engines due to their simultaneous reduction of smoke and NO_x emissions using vegetable oils/diesel as fuel [11–14]. Emulsification is a simple process and needs no modification in the engine design. Emulsion properties are further improved by mixing alcohols with oil and water during the emulsion preparation process. Alcohols act as cosurfactants and fuel extenders in making emulsions. They increase the stability and reduce viscosity of emulsions [15–19]. Since animal fats mix freely with alcohols, alcohols can be used as cosurfactants to improve the emulsion properties further. Past results obtained from the experiments on diesel engines using emulsions of animal fats/diesel with alcohol showed very good agreement with the engine performance and exhaust smoke emissions [15,16,20].

In this work, different methods, such as neat fat operation, fat preheating, and emulsions with methanol/ethanol, are studied experimentally to analyze the emissions and combustion characteristics of a diesel engine fueled with animal fat. Experiments are conducted with different fractions of alcohols, water, and fats.

---

[1] To whom correspondence should be addressed.

Contributed by the Internal Combustion Division of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received October 4, 2004; final manuscript received October 17, 2005. Review conducted by J. C. Cowart.



1. Test Engine
2. Dynamometer
3. Animal Fat Tank
4. Diesel Tank
5. A/D Card for Pressure
6. A/D Card for Analyser
7. Air Tank
8. Burette for diesel
9. Burette for Animal Fat
10. Charge Amplifier
11. Fast Data Acquisition Sysytem
12. Slow Data Acquisition system
13. Cylinder Pressure Sensor
14. Injection Pressure Sensor
15. Diesel Filter
16. Animal fat Filter
17. TDC shaft position Encoder
18. Speed Sensor
19 Exhaust gas Analyser
20. Smoke Meter

Fig. 1 Schematic of experimental setup

Fuel inlet temperature is also varied at different temperatures for neat fat operation. Results are optimized based on minimum emission levels. The optimum results obtained with different methods are compared to their respective fats and neat diesel at two different power outputs (i.e., 100% load and 60% load). Comparison is made for the following cases:

(a) neat diesel and neat animal fat at normal temperature
(b) animal fat with preheating at 70°C and emulsions of animal fat with methanol
(c) neat animal fat and emulsions of animal fat with ethanol

Experimental Setup and Experimental Procedure

Engine Test Cell. A single-cylinder air-cooled Lister Petter (TS1) diesel engine developing a power output of 2.8 kW at 1500 rpm is used for the work. The schematic of the experimental setup is shown in Fig. 1, and the engine details are given in Table 1. An electrical dynamometer is used for loading the engine. An orifice meter connected to a large tank is attached to the engine intake manifold to make airflow measurements. The fuel flow rate is measured on the volumetric basis using a burette and a stopwatch. A chromel-alumel thermocouple in conjunction with a slow-speed digital data acquisition system is used for measuring the exhaust gas temperature.

Table 1 Engine details

Make	Lister Petter—TS 1
General details	Four-stroke, compression ignition, air-cooled, naturally aspirated, single-cylinder engine
Bore and stroke	95.3 mm × 88.9 mm
Connecting rod length	165.3 mm
Compression ratio	18:1
Rated brake power output	2.8 kW at 1500 rpm
Injector opening pressure	250 bar
Displacement volume	630 cc
Fuel injection timing	20 deg BTDC (static)

Table 2 Properties of diesel, animal fat, and vegetable oil

Properties	Diesel	Animal fat	Vegetable oil
Density (kg/m^3)	840	920	904
Lower heating value (kJ/kg)	42,490	39,770	37,000
Viscosity ($\times 10^{-6}$ m^2/s) at 30°C	4.59	45	40
Carbon (% by mass)	84–87	73	77.6
Hydrogen (% by mass)	13–16	12.3	11.6
Oxygen (% by mass)	00	12.5	10.8
Sulphur (% by mass)	0.29	0.0	0.0
Fatty acid composition [21,22]			
Oleic acid (% by mass)		42.1	64.1
Palmitic acid (% by mass)		22.7	3.5
Linoleic acid (% by mass)		17.1	22.3
Palmitoleic acid (% by mass)		8.3	0.1
Stearic acid (% by mass)		5.4	0.9

Emission Instrumentation. An infrared (COSMA) exhaust analyzer is used for measuring hydrocarbon (HC) and carbon monoxide (CO) emissions. NO in the exhaust is measured by using a Beckman chemiluminascence NO/NOx analyzer. Smoke levels are measured by using a standard Hartridge smoke meter, which works on a light absorption technique (passing a light beam through the exhaust sample and the fraction of light is absorbed by the exhaust gas). Light extinction coefficient K is used as the measure of smoke density. Details on smoke measurement can be found in [18,20].

Combustion Data Acquisition. A high-speed digital data acquisition system in connection with AVL 620-Indiwin hardware is used for combustion data acquisition. The AVL INDISET 620 consists of a docking station with 16 input channels connected to a PC via a centronics interface. The combustion pressure is measured using a piezoelectric transducer (AVL Model) mounted flush on the cylinder head. The fuel line pressure is measured by an another piezoelectric transducer (AVL Model) mounted on the fuel line very close to the fuel injector. Both transducers are connected with the high-speed data acquisition system to obtain cylinder pressure and fuel-line pressure histories. The output signals generated by the transducers are conditioned by the appropriate charge amplifiers. An AVL (Model 364) crankshaft position encoder is used to give signals at TDC. Engine in-cylinder pressure and crank-angle signals are sampled for 100 consecutive cycles at the increments of 0.1 crank-angle intervals.

Experimental Procedure. Animal fats of the same kind (i.e., duck fat) are collected from several fat industries. The fats used for making emulsions are obtained by centrifuging, heating, and separating the raw waste fat. Before conducting all the experiments, preliminary analysis is performed on the animal fat to obtain important properties, such as viscosity, density, lower heating value, etc., to find its suitability as fuel for diesel engines. Fatty acid compositions are found from the literature for animal fats [21,22]. The obtained properties of animal fat are compared to diesel and vegetable oil in Table 2. (Further details on animal fats properties and their measurement can be found in [19]). Methanol animal fat emulsions are then prepared in the laboratory by adding 2% of surfactant (Span 83 also called as sorbitan sesquiolate) by volume to animal fat and thereafter adding a water/methanol mixture to the animal fat. Span 83 is chosen as the surfactant for all formulations because of the better stability of emulsions. Some of the properties can be seen in Table 3. The same procedure is followed for ethanol animal fat emulsions also using different proportions of fat, surfactant, water, and ethanol. A number of formulations are made by varying water, surfactant, alcohol, and fat fractions to obtain the optimum formulation [19,23]. It was found that the methanol and ethanol emulsions were stable up to 15 days without any phase separation.

Experiments are then carried out on the engine using diesel and neat fat as fuels. During the entire investigation, the injection

Table 3 Properties of surfactant—Span 83

Chemical name	Sorbitan sesquiolate
Molecular formula	$C_{66}H_{108}O_{13}$
HLB number	3.7
Molecular weight	1110
Fatty acid composition	Oleic acid 70%, balance primarily palmitic acid, strearic acid and linoleic acid.
Vapor pressure	0.81 psi at 20°C
Density	0.989 g/ml at 25°C

timing is optimized and set at 20 deg before top dead center (TDC) based on minimum emission levels. The engine is thermally stabilized before taking all measurements. Readings of engine speed, fuel flow, air flow, exhaust gas temperature, etc., are recorded for obtaining performance parameters. Exhaust gas analyzers are calibrated carefully before making measurements, based on the manufacturer's recommended procedure. Standard span gases and zero gas are used for the calibration of HC, CO, and NO. Observations are made for smoke, NO, HC, and CO to analyze the emission characteristics. In all cases, pressure crank angle data are recorded and processed to get combustion parameters (such as cylinder peak pressure, ignition delay, combustion duration, and heat release rate). Furthermore, tests are repeated with preheated animal fat prior to injection at different temperatures (40°C, 50°C, 60°C, and 70°C) to obtain the optimum temperature for minimum emissions. In the next phase, experiments are carried out with methanol animal fat emulsions with different fractions of water, surfactant, and methanol using animal fat. Finally, experiments are done with ethanol animal fat emulsions with different fractions of water, surfactant, ethanol, and animal fat. The optimum formulations among them are found based on minimum exhaust emission levels in all cases. Detailed analysis can be found in the previous study [18,20]. Optimum formulations with different emulsions can be seen in Table 4. Optimum results of the emulsions of ethanol and methanol and fuel preheating (i.e., 70°C) are compared to their respective neat fats and diesel at two different power output conditions (i.e., 60% and 100% load).

Results and Discusssion

Combustion Parameters. Cylinder pressure crank-angle variations obtained by averaging 100 cycles at peak power output with different methods tested are given in Figs. 2 and 3. All the tested fuels follow the trend, similar to the diesel pressure diagram. The cylinder peak pressure is highest with diesel followed by animal fat emulsion and the neat animal fat as seen in Fig. 2. The same trend is observed in case of fat preheating also in Fig. 3. However, animal fat emulsion shows a small deviation in occurrence of peak pressure as compared to neat fat and neat diesel. The delayed start of combustion and resulting increase in peak pressure over that for neat animal fat due to the strong premixed combustion phase (will be explained later) are clearly seen in Fig. 2. It can be noted that the occurrence of peak pressure moves further away from top dead center for the emulsion in comparison to neat animal fat and diesel fuels. This indicates that the ignition delay (which will be explained later) is longer with the emulsions as compared to neat fat.

Table 4 Formulation of best emulsions

Best emulsion	Water fraction (%)	Animal fat fraction (%)	Alcohol fraction (%)	Surfactant Fraction (%)
Methanol animal fat best emulsion	9.8	78.4	9.8	2
Ethanol animal fat best emulsion	10	50	36	4

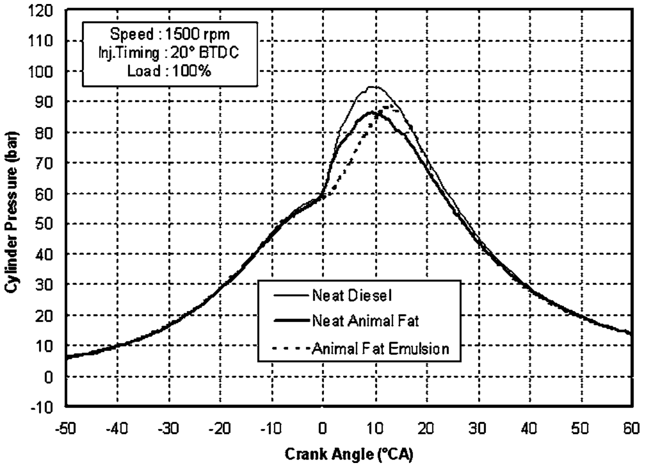


Fig. 2 Cylinder pressure crank-angle diagram with methanol animal fat emulsion at maximum power output

A comparison of cylinder peak pressure and maximum rate of pressure rise at peak and part (i.e., 100% and 60% load) power outputs with different methods are shown in Figs. 4 and 5. Neat animal fats result in lower peak pressure and rate of pressure rise as compared to neat diesel fuel at normal temperature. In a com-

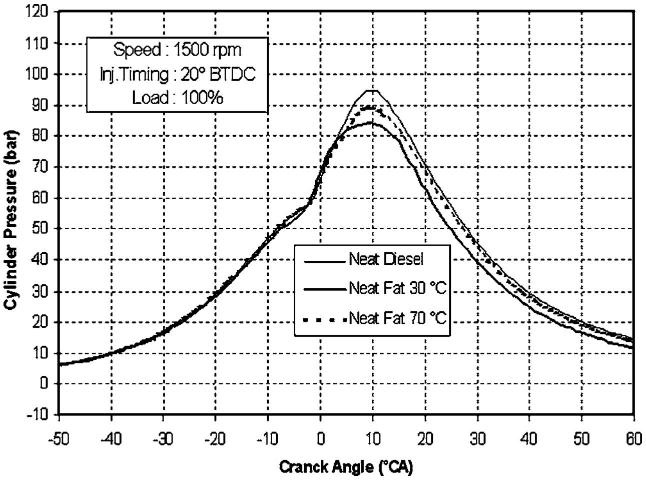


Fig. 3 Cylinder pressure crank-angle diagram with preheated fat at maximum power output

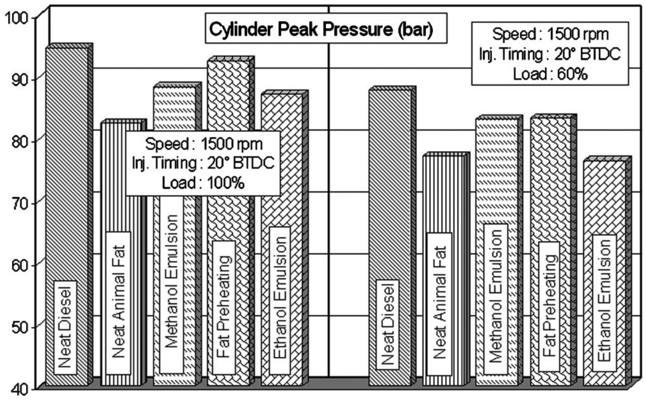


Fig. 4 Variation of cylinder peak pressure with different methods

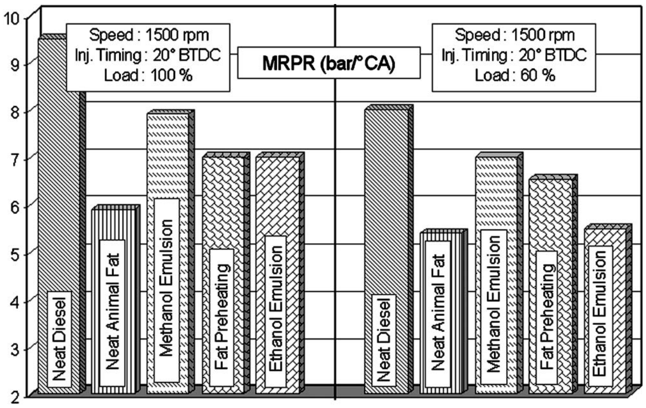


Fig. 5 Variation of maximum rate of pressure rise with different methods

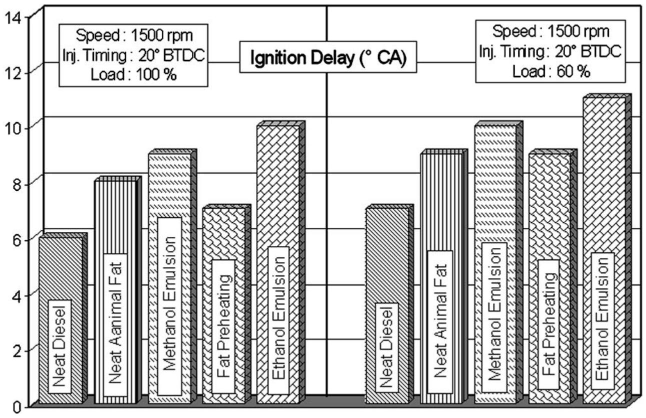


Fig. 6 Variation of ignition delay with different methods

pression ignition engine, the peak pressure depends on the combustion rate in the initial stages, which, in turn, is influenced by the fuel taking part in the uncontrolled combustion. The uncontrolled or the premixed combustion phase is governed by the delay period, the spray envelope, and the air-fuel mixture preparation during the delay period. Thus, the higher viscosity and poor volatility of the neat animal fat at normal temperature result in lower peak pressure and maximum rate of pressure rise as compared to neat diesel. However, there is an improvement in peak pressure and maximum rate of pressure rise with the preheated animal fat and the emulsions of animal fat with methanol and ethanol. The maximum cylinder pressures at peak power output are found as 95 bar, 82 bar, 93 bar, 88 bar, and 87 bar, respectively, with neat diesel, neat fat, preheated animal fat, methanol animal fat emulsion, and ethanol animal fat emulsion. The increase in peak pressure with the emulsions is due to the enhanced combustion rate as a result of rapid combustion of emulsions at the premixed combustion period. The dispersed water droplets in the evaporating spray have much lower boiling temperatures than the surrounding fuel. Under certain conditions they become superheated and, subsequently, expand in a very rapid vaporization event called microexplosion. Presence of water and methanol fractions also lowers the temperature of the combustion chamber and leads to more fuel being accumulated during the ignition delay period. A strong premixed combustion rate due to a long ignition delay results in higher peak pressure and rate of pressure rise as compared to neat fat. With the preheated animal fat, vaporization of the fat becomes better because of the improved viscosity and combustion becomes faster due to rapid burning of the injected fuel.

The variation of ignition delay with all the methods is shown in Fig. 6. The ignition delay period of all the fuels tested is calculated based on the dynamic injection timing. The duration between the point of the start of injection to the point of ignition is taken as the ignition delay. The point of fuel injection is found by using a piezoelectric pressure sensor that gives the online fuel injection pressure. The start of combustion is determined from the rate of pressure rise variation. This shows a sudden rise in the slope at the point of ignition due to the high premixed heat release rate. Ignition delay shown in Fig. 6 is longer with neat animal fat as compared to neat diesel due to the low cetane number. With neat animal fat, due to poor atomization and vaporization, physical delay becomes longer as compared to neat diesel. The ignition delay is found as 6 deg CA (crank angle) with neat diesel and 8 deg CA with neat animal fat at normal temperature. Ignition delay further increases with both animal fat emulsions as compared to neat animal fat and neat diesel fuel. It is found as 9 deg CA with methanol animal fat emulsion and 10 deg CA with ethanol animal fat emulsion at peak power output. The increase in ignition delay with animal fat emulsions is due to the high latent heat of vaporization of water and methanol/ethanol in the emulsions. Vaporization of water and alcohols reduces the temperature of intake air and fuel. In addition, the presence of water and methanol/ethanol results in reduction of the overall cetane number of the emulsions. Hence, the delay is longer with the emulsions as compared to neat fat. However, ignition delay reduces with fat preheating. It is found as 7 deg CA at peak power output.

Figure 7 shows the variation of combustion duration. The combustion duration is calculated by obtaining the cumulative heat release rate. The end of combustion is taken as the point where 95% of the heat release had occurred. Longer combustion duration is observed with neat fats than diesel at both power outputs. This is due to the injection of larger quantities of animal fat than diesel for the same load condition (because the heating value of animal fat is lower than diesel fuel). Since the diffusion burning (will be explained later) is more pronounced with the animal fat, late burning occurs in the expansion stroke and results in longer combustion duration with the neat fats. However, the combustion duration is reduced with both the emulsions of animal fat as compared to neat fat. Because of the long ignition delay, more fuel is physically prepared (evaporation, mixing, etc.) with the emulsions for chemical reaction, and rapid burning occurs in the premixed stage itself. Hence, the heat release during the diffusion burning period is lowered and results in reduced combustion duration. The microexplosion further accelerates diffusion combustion and decreases total combustion duration. Preheating also indicates reduction in combustion duration as compared to neat fat at both power outputs.

Heat release patterns with neat fat operation and other tested methods are compared in Figs. 8–11. The heat release rate is

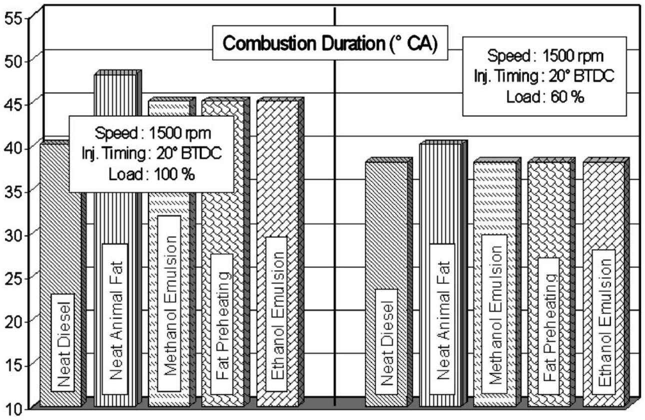


Fig. 7 Variation of combustion duration with different methods

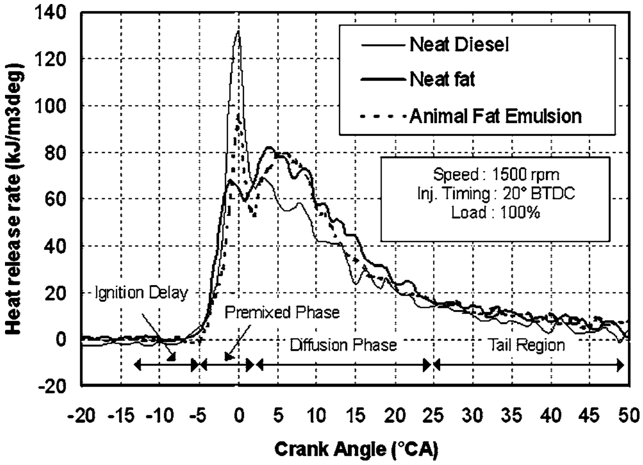


Fig. 8 Variation of heat release rate with methanol emulsion at peak power output

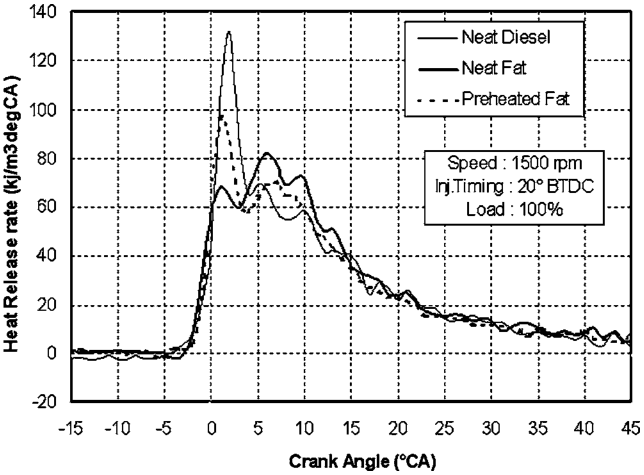


Fig. 10 Variation of heat release rate with preheated animal fat at peak power output

calculated by performing the first law analysis of the average pressure versus crank-angle variations. As in thermodynamics, a simplified calculation process that determines the energy effectively delivered to the gas is taken in to account. Fuel vapor and products are treated as the mixture of ideal gases. The surface heat transfer losses are ignored. The gas mass is taken for determination of the gas temperature and the heat release equation is written as

$$dQ = \left(\frac{\gamma}{\gamma - 1}\right) PdV + \left(\frac{1}{\gamma - 1}\right) VdP \tag{1}$$

where γ is the ratio of specific heats, P is the cylinder pressure and V is the instantaneous volume.

From the Figs. 8–10, it is seen that the premixed burning is more pronounced with diesel as expected. Neat animal fat shows lower heat release rate at the initial stage as compared to neat diesel. The high viscosity and density of neat animal fat result in poor atomization and vaporization and lead to reduction in air entrainment and fuel-air mixing rates. Hence, more burning occurs in the diffusion phase. With fat preheating, there is an improvement in heat release rate as seen in Fig. 10. By raising the temperature, the premixed phase of the heat release curve becomes high due to the improved atomization and vaporization of the animal fat. The low viscosity of the preheated fat leads to form more flammable fuel-air mixture during the delay period and enhances the combustion process. This results in improved heat release rates.

Animal fat emulsions with methanol and ethanol also show improvement in heat release rates (both in premixed and diffusion combustion) as compared to neat fat as shown in Figs. 8 and 9. There is a delay in the start of combustion and an increase in the heat release rate at the premixed burn period with the emulsion as compared to neat animal fat. The diffusion combustion phase is less with the emulsions as compared to neat fat. When the ignition delay is increased, more fuel is accumulated in the combustion chamber and physically prepared for chemical reaction. Once the accumulated fuel attains its self-ignition temperature, it burns instantaneously and raises the peak pressure and the premixed combustion rate. In addition, microexplosion of droplets enhances the combustion rate. The good atomization and vaporization of emulsions promote rapid mixing with the surrounding air [24]. The oxygen available in the fuel further improves the overall rate of combustion. All these factors contribute to the improved heat release rates with the emulsions as compared to neat fat. However, at part load the improvement in heat release rate is not much significant (Fig. 11).

Performance Parameters. Neat animal fat as indicated in Fig. 12 results in increased specific energy consumption (SEC) as

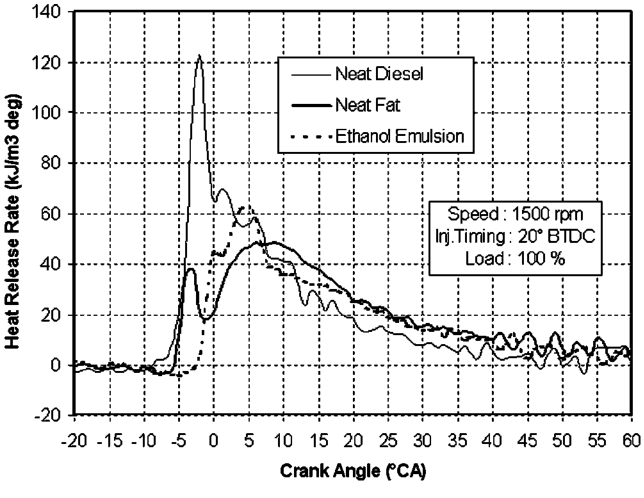


Fig. 9 Variation of heat release rate with ethanol emulsion at peak power output

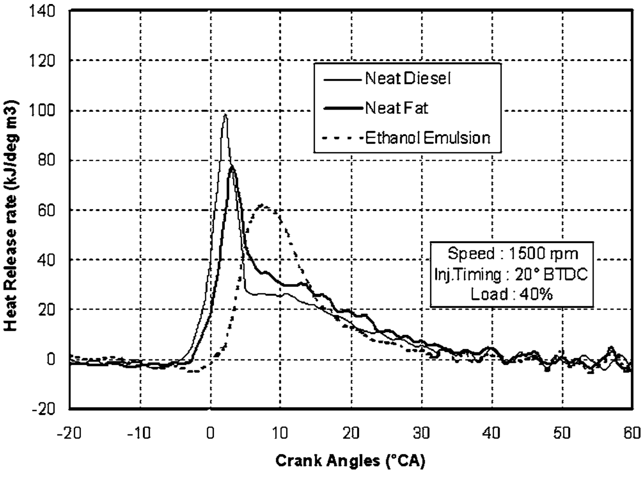


Fig. 11 Variation of heat release rate with ethanol emulsion at part load

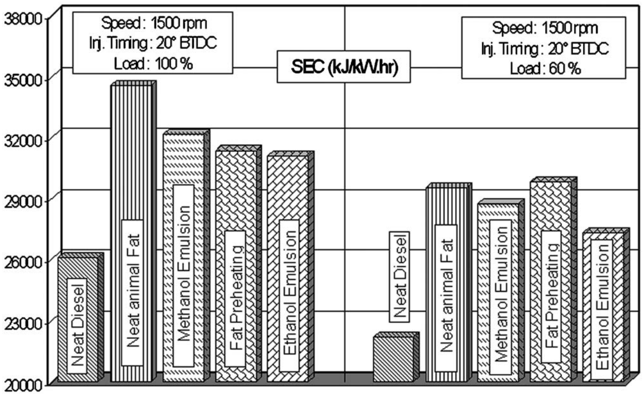


Fig. 12 Variation of specific energy consumption with different methods

compared to neat diesel at 100% and 60% of the maximum engine power output. High viscosity and poor volatility of the fats result in poor atomization and mixture formation and in higher specific energy consumption than neat diesel. However, there is an improvement in specific energy consumption with all the methods adopted. The reduction in SEC with fat preheating is due to the high combustion rate. The preheated fat has lower viscosity, which results in better atomization of the fuel as compared to neat fat at normal temperature. Emulsions of animal fat result in reduced SEC due to the better fuel atomization. Microexplosion of the emulsions leads to secondary atomization [25] and reduces the mean diameter of the injected fuel. In addition, the presence of surfactant in the emulsions contributes to better atomization of the fuel. Ethanol emulsion reaches the minimum SEC as compared to methanol due to its superior physical characteristics, such as low viscosity and better miscibility with animal fat than methanol, and results in overall reduction in viscosity of emulsions [23]. However, SEC is higher with all the methods as compared to neat diesel. The difference becomes more at 60% load. At part loads, due to heat loss to the walls, SEC becomes high despite the fact that flame temperature is not as high in the case of emulsion (cf. NO_x analysis). This observation is not dramatic because the aim is to obtain a fuel for CI engines from a waste that is nearly free.

The exhaust gas temperature as shown in Fig. 13 is very high with neat animal fat as compared to neat diesel due to slow combustion. With the emulsions there is a reduction in exhaust gas temperature. Ethanol emulsion shows the highest reduction in exhaust gas temperature (i.e., 480 °C at the optimum emulsion, where as it is 580 °C with its neat animal fat. This reduction in exhaust gas temperature is due to the reduction in charge temperature as a result of vaporization of ethanol. A similar trend is seen in the case of methanol emulsion also. The high latent heat of vaporization of alcohols and water helps in reducing the cylinder temperature with both the emulsions. However, with fat preheating due to the high fuel inlet temperature, exhaust gas temperature becomes high mainly at high power output.

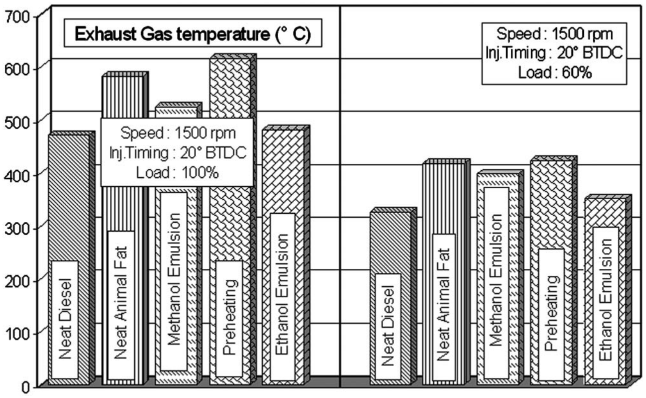


Fig. 13 Variation of exhaust gas temperature with different methods

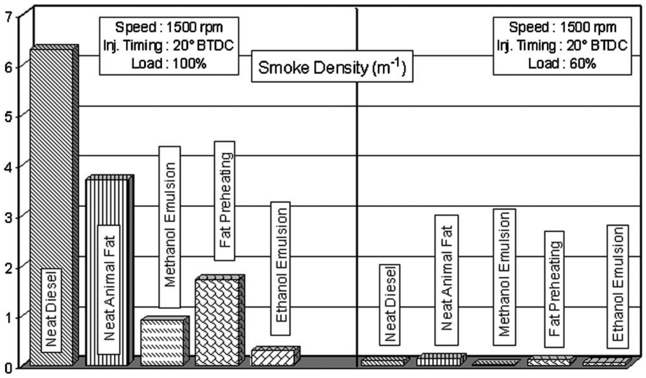


Fig. 14 Variation of smoke density with different methods

Emission Parameters. The smoke level is indicated in Fig. 14. It is interesting to see that neat animal fat result in lower smoke levels than neat diesel mainly at peak power output. It is about 3.7 m^{-1} with neat animal fat and 6.3 m^{-1} with neat diesel at the maximum power output. This reduction in smoke emission with neat fats is due to the presence of in-built oxygen with the fats. The high oxygen content in the animal fats helps in complete oxidation of the fuel and reduces soot concentration in the exhaust gas. Smoke further reduces with fat preheating and emulsions of methanol and ethanol. The values are found as 1.7 m^{-1} with preheated animal fat, 0.89 m^{-1} with methanol emulsion, and 0.28 m^{-1} with ethanol emulsions. Improved vaporization of the preheated fat results in lower smoke values than fat at normal temperature. Emulsions of animal fat promote this reduction due to the presence of alcohol content. Microexplosion plays a major role in drastic reduction in smoke emissions [11,13]. It leads to secondary atomization and permits one to obtain a better fuel-air mixture formation. Hence, smoke density reaches to very low values with animal fat emulsions. The greatest reduction in smoke emission is seen with ethanol animal fat emulsion than with methanol and fat preheating. The combustion of ethanol animal fat emulsion produces the minimum smoke emission because of higher ethanol content as compared to methanol. However, at part load, the smoke levels are very low with all the fuels and the differences are not distinguishable.

Shown in Fig. 15 is the variation of hydrocarbon emission with different methods at 100% and 60% load conditions. The hydrocarbon emission at normal temperature is higher with neat animal fat as compared to neat diesel at both power outputs. Unburned hydrocarbons are the results of incomplete combustion. High viscosity and poor volatility of animal fat result in poor mixing of the fuel with air and lead to more hydrocarbon emissions at normal temperature. However with the preheated fat, there is a reduction in hydrocarbon emissions. It can be noted that the HC emission with animal fat approaches diesel value at a high fuel inlet temperature of 70 °C. Because of the improved vaporization and fuel-air mixing rates, combustion becomes complete and results in low hydrocarbon emissions with the preheated animal fat. Emulsions of animal fat also indicate lower levels of hydrocarbon emissions as compared to their parent fuels mainly at 100% power output. This is due to the reduction in the overall amount of carbon admitted into the engine. Secondary atomization provided by the microexplosion of water droplets increases the surface area of contact of fuel droplets with air, improves the fuel-air mixture formation, and leads to lower hydrocarbon emissions as compared

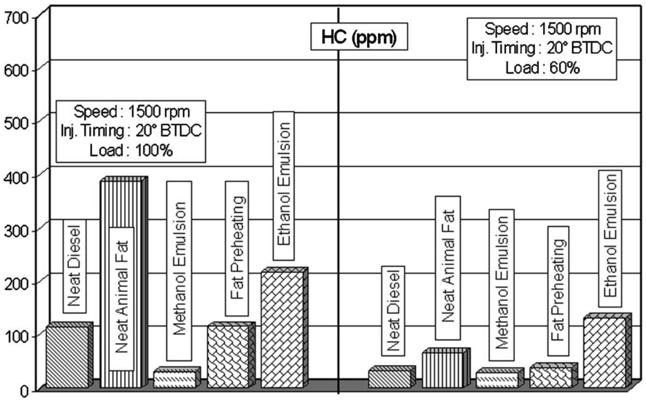


Fig. 15 Variation of hydrocarbon emissions with different methods

to neat fat. However, ethanol emulsion shows sight increase in HC than neat animal fat at light load (i.e., 60% of maximum power output). This finds explanation that the larger amount ethanol present in the emulsion causes lower combustion temperatures and leads to partial combustion of the fuel at part loads.

Neat animal fat leads to higher CO emissions than diesel at normal temperature as shown in Fig. 16. As mentioned earlier, fuel richness due to low volumetric efficiency contributes to the trend of higher CO emissions with neat fats. Rich pockets formed in the cylinder cause more CO emissions with animal fat at normal temperature. It may be noted that the lower heating value of animal fat leads to injection of higher quantities of fuel as compared to diesel for the same load conditions. However, fuel preheating leads to complete combustion of the fuel and reduces CO emission. The level becomes lower than diesel with preheated animal fat at 70 °C at both power outputs. Emulsions of animal fat with methanol and ethanol show significant reduction in CO emissions as compared to their respective fats. Introduction of water into the fuel replaces a portion of flammable fuel that contains hydrocarbons. The increase in vaporized fuel jet momentum gives great air entrainment to the fuel jet and accelerates the diffusive burning rate. In addition, the presence of water and methanol in the emulsion increases oxygen concentration in the fuel, which helps in complete oxidation of the fuel. All the factors participate in significant reduction in CO emissions. It is interesting to note that the levels with the emulsions are even lower than neat diesel operation. The trend of CO emissions is also similar for all the methods in part load.

The variation of NO emission with power output is shown in Fig. 17. NO formation in diesel engine is due to the high combustion temperature and the availability of oxygen. It forms mainly in the high-temperature regions of the product gases. It is seen that the neat animal fat emits lower NO levels as compared to standard diesel at both power outputs. The reduction in NO emission with animal fat is mainly associated with the reduced premixed burning rate following the delay period. The lower air entrainment and fuel-air mixing rates with the animal fat result in low peak temperature and the resulting NO levels. Fuel preheating shows a rising trend in NO emissions due to rapid burning as a result of increased fuel inlet temperatures. This is the drawback with fuel preheating. However, the values are still lower than diesel. NO further reduces with the emulsions with methanol and ethanol. This reduction in NO emission is due to the drop in charge temperature as a result of vaporization of alcohol and water. The water brought in by the emulsified fuel changes the relative quantities of fuel, oxygen, and inert during the rich premixed-burn stage of animal fat combustion. Water lowers the peak combustion temperatures due to its high latent heat of vaporization. This reduces formation of nitrogen oxides. It must be noted that the NO level is very low, even with the increased premixed combustion phase in case of emulsions. In general, with animal fat emulsions the NO emissions decrease considerably at both power outputs without compromising engine performance.

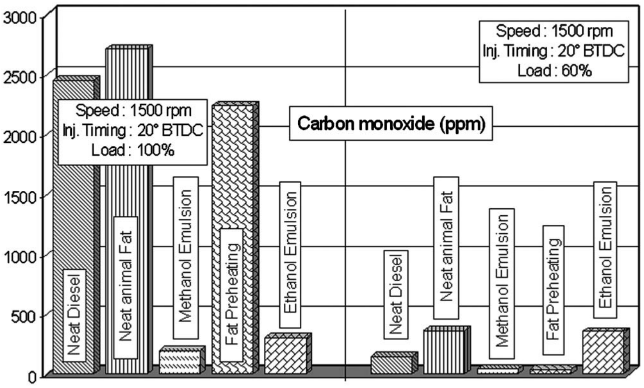


Fig. 16 Variation of carbon monoxide emissions with different methods

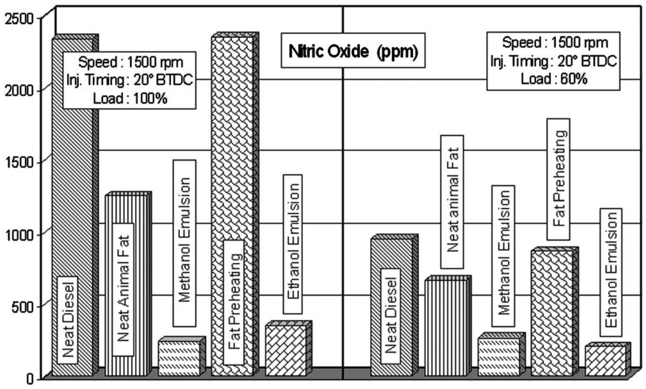


Fig. 17 Variation of nitric oxide emission with different methods

Conclusion

The following conclusions are made based on the above results:

- Neat animal fat results in lower cylinder peak pressure, maximum rate of pressure rise, longer ignition delay, and combustion duration as compared to neat diesel at 60% and 100% power outputs. Increased specific energy consumption, exhaust gas temperature, hydrocarbon, and carbon monoxide emissions are observed with neat fat as compared to neat diesel at both power output conditions. Lower smoke levels are found with neat fat at both power outputs due the in-built oxygen present in it. NO emissions are also found as lower with neat fat as compared to neat diesel due to slow combustion.
- Emulsions of animal fat with methanol/ethanol and fat preheating show higher peak pressure and maximum rate of pressure rise as compared to neat fat at both power outputs. But the values are still lower than diesel. Longer ignition delay and combustion duration are observed with neat animal fat as compared to neat diesel. Emulsions of animal fat further prolong the ignition delay. However, combustion duration is shorter with the emulsions. Preheating shows shorter ignition delay and combustion duration as compared to neat fat. The heat release rate shows improvement with all the methods adopted as compared to neat animal fat.
- Emulsions of animal fat with alcohols show considerable

Table 5 Summary of the results obtained with different fuels tested and their effects relative to base diesel at peak power output

Methods	Neat fuel operation		Emulsified fuel	
Fuels	Neat fat	Preheated fat	Methanol	Ethanol
SEC	↑	↑	↑	↑
EGT	↑	↑	↑	↑
Smoke	↓	↓	↓	↓
NO	↓	↑	↓	↓
CO	↑	↓	↓	↓
HC	↑	↑	↓	↓
PP	↓	↓	↓	↓
MPPR	↓	↓	↓	↓
ID	↑	↑	↑	↑
CD	↑	↑	↑	↑
HRR	↓	↑	↑	↑

improvement in SEC and reduction in exhaust gas temperature, hydrocarbon, and carbon monoxide emissions as compared to their parent fuels mainly due to microexplosion. Preheating also shows lower SEC, hydrocarbon, and carbon monoxide emissions. However, the exhaust temperature is higher with preheating. Animal fat emulsions (both methanol and ethanol) and fat preheating show considerable reduction in smoke density as compared to their neat fats. However, there is no significant difference in smoke emissions with all the fuels at low power output. Emulsions further reduce NO emissions due to the high latent heat of vaporization of water and alcohols.

From the above results it can be concluded that preheated animal fat and emulsions of animal fat with methanol/ethanol can be used as fuel in a diesel engine with improved performance and reduced emissions as compared to neat fat. Emulsification of animal fat with methanol and/or ethanol can be preferred as better methods to use animal fat efficiently in a diesel engine with a drastic reduction in all emissions as compared to fat preheating. Preheating can also lead to a slight improvement in engine performance and emissions without modifying the fuel. However, measures must be taken to control NO emissions with fat preheating. Table 5 presents a summary of the results obtained with different methods using animal fat as base fuel and their effects relative to neat diesel. Stability of emulsions and the longterm effects of emulsions on engine parts need further study.

Nomenclature

BTDC = before top dead center
CA = crank angle
CO = carbon monoxide
DDAS = digital data acquisition system
HC = Hydrocarbon
P = cylinder pressure
V = cylinder volume
Q = heat release
γ = ratio of specific heats

References

[1] Barsic, N. J., and Humke, A. L., 1981, "Performance and Emission Characteristics of a Naturally Aspirated Diesel Engine With Vegetable Oils," Society of Automotive Engineers Paper No. 810262.
[2] Vellguth, G., 1983, "Performance of Vegetable Oils and Their Monoesters as Fuel for Diesel Engines," Society of Automotive Engineers Paper No. 831358.
[3] Senthil Kumar, M., Ramesh, A., and Nagalingam, B., 2001, "Complete Vegetable Oil Fuelled Compression Ignition Engine," Society of Automotive Engineers Paper No. 2001-28-0067.
[4] Senthil Kumar, M., Kerihuel, A., Bellettre, J., and Tazerout, M., 2005, "Investigations on the Use of Preheated Animal Fat as Fuel in a Diesel Engine," Renewable Energy, **30**(9), pp. 1443–1456.
[5] Senthil Kumar, M., Ramesh, A., and Nagalingam, B., 2000, "Performance Studies on a CI Engine Using Methyl Ester of Jatropha oil as Fuel," *Proceedings of XVI National Conference on I. C. Engines and Combustion*, Calcutta, pp. 89–94.
[6] Karim, G. A., 1987, "The Dual Fuel Engine," *Automotive Engine Alternatives*, R. L. Evens, Bleham Press.
[7] Senthil Kumar, M., Ramesh, A., and Nagalingam, B., 2001, "Experimental Investigations on a Jatropha Oil Methanol Dual Fuel Engine," *New Developments in Alternative Fuels for IC Engines* (SP-1608), Society of Automotive Engineers, Warrendale, PA, SAE Paper No. 2001-01-0153.
[8] Senthil Kumar, M., Ramesh, A., and Nagalingam, B., 2003, "An Experimental Comparison of Methods to Use Methanol and Jatropha Oil in a Compression Ignition Engine," Biomass Bioenergy, **25**(3), pp. 309–318.
[9] Nwafor, O. M. I., 2003, "The Effect of Elevated Fuel Inlet Temperature on Performance of Diesel Engine Running on the Neat Vegetable Oil at Constant Speed Conditions," Renewable Energy, **28**, pp. 171–181.
[10] Bari, S., Lim, T. H., and Yu, C. W., 2002, "Effect of Preheating of Crude Palm Oil (CPO) on Injection System, Performance and Emission of a Diesel Engine," Renewable Energy, **27**, pp. 339–351.
[11] Marwan, A. A. N., Hobina, R., and Wagstaff, S. A., 2001, "The Use of Emulsion, Water Induction and Egr for Controlling Diesel Engine Emissions," Society of Automotive Engineers Paper No. 2001-01-1941.
[12] Yoshimoto, Y., Onodera, M., and Tamaki, H., 1999, "Reduction of NOx, Smoke and BSFC in a Diesel Engine Fuelled by Biodiesel Emulsion With Used Frying Oil," Society of Automotive Engineers Paper No. 999-01-3598.
[13] Hsu, B. D., 1986, "Combustion of Water-in-Diesel Emulsion in an Experimental Medium Speed Diesel Engine," Society of Automotive Engineers Paper No. 860300.
[14] Lin, C. Y., and Wang, K. H., 2004, "Diesel Engine Performance and Emission Characteristics Using Three-Phase Emulsions as Fuel," Fuel, **83**, pp. 537–545.
[15] Baker, Q. A., 1981, "Use of Alcohol-in-Diesel Fuel Emulsions and Solutions in a Medium Speed Diesel Engine," Society of Automotive Engineers Paper No. 810254.
[16] Haepp, H. J., and Truong, H. S., 1983, "The Effect of Methanol/Diesel Fuel Emulsions on the Mixture Formation in Direct Injection Diesel Engines. A Theory on Spontaneous Evaporation," Society of Automotive Engineers Paper No. 830376.
[17] Khan, N., and Gollahalli, S. R., 1981, "Performance and Emission Characteristics of a Diesel Engine Burning Unstabilized Emulsions of Diesel Fuel With Water Methanol and Ethanol," Society of Automotive Engineers Paper No. 811210.
[18] Kerihuel, A., Senthil Kumar, M., Bellettre, J., and Tazerout, M., 2005, "Experimental Investigations on a Compression Ignition Engine Using Animal Fat Emulsions as Fuel With Water and Methanol," Society of Automotive Engineers Paper No. 2005-01-1729.
[19] Kerihuel, A., Senthil Kumar, M., Bellettre, J., and Tazerout, M., 2005, "Use of Animal Fats as CI Engine Fuel by Making Stable Emulsions With Water and Methanol," Fuel, **84**, pp. 1713–1716.
[20] Senthil Kumar, M., Kerihuel, A., Bellettre, J., and Tazerout, M., 2005, "Effect of Water/Methanol Fractions on the Performance of a Compression Ignition Engine," Asph. Paving Technol. (to appear).
[21] Goodrum, J. W., Geller, D. P., and Adams, T. T., 2003, "Rheological Characterization of Animal Fats and Their Mixtures With # 2 Fuel Oil," Biomass Bioenergy, **24**, pp. 249–256.
[22] Demirba, A., 2002, "Biodiesel From Vegetable Oils via Transesterification in Supercritical Methanol," Energy Convers. Manage., **43**, pp. 2349–2356.
[23] Kerihuel, A., Senthil Kumar, M., Bellettre, J., and Tazerout, M., 2005, "Ethanol Animal Fat Emulsions as a Diesel Engine Fuel—Formulations and Influential Parameters (P. 1)," Fuel (submitted).
[24] Musculus, M. P. B., Dec, J. E., Tree, D. R., Daly, D., Langer, D., Ryan, T. W., and Matheaus, A. C., 2002, "Effects of Water-Fuel Emulsions on Spray and Combustion Processes in a Heavy-Duty DI Diesel Engine," Society of Automotive Engineers Paper No. 02012892.
[25] Adiga, K. C., and Shah, D. O., 1990, "On the Vaporization Behavior of Water-in-Oil Microemulsions," Combust. Flame, **80**, pp. 412–414.

Ali Mohammadi
Toyota Motor Europe,
Hoge Wei 33B-1930 Zaventem, Belgium
e-mail: ali.mohammadi@toyota-europe.com

Masahiro Shioji
e-mail: shioji@energy.kyoto-u.ac.jp

Takuji Ishiyama

Masato Kitazaki

Graduate School of Energy Science,
Kyoto University,
Sakyo-ku, Kyoto 606-8501, Japan

Utilization of Low-Calorific Gaseous Fuel in a Direct-Injection Diesel Engine

Low-calorific gases with a small portion of hydrogen are produced in various chemical processes, such as gasification of solid wastes or biomass. The aim of this study is to clarify the efficient usage of these gases in diesel engines used for power generation. Effects of amount and composition of low-calorific gases on diesel engine performance and exhaust emissions were experimentally investigated adding hydrogen-nitrogen mixtures into the intake gas of a single-cylinder direct-injection diesel engine. The results indicate that optimal usage of low-calorific gases improves NO_x *and Smoke emissions with remarkable saving in diesel fuel consumption.* [DOI: 10.1115/1.2179464]

Introduction

Both municipal solid waste generation and energy consumption are increasing in developed and developing countries due to growth in the world's population and living standards [1,2]. Efficient energy generation and proper waste management are necessary to maintain living standards and to protect our environment. Thermal gasification is a technology that could contribute to power generation in the management of municipal waste and utilization of biomass [3,4]. Improved energy extraction and pollution control would make it more attractive than conventional waste treatment, typically dominated by incineration and land filling. The literature shows that waste from commercial/residential buildings and factories can be converted to low-calorific gaseous fuels employing decentral and small-scale high-temperature steam/air reforming system [5–7]. Low-calorific gas (LCG) contains mainly nitrogen gas with a small fraction of hydrogen and carbon monoxide. Depending on chemical composition, the heating value of LCG varies in the range of 3500–5000 $\mathrm{kJ/Nm^3}$. This corresponds roughly to 10% of the calorific value of natural gas. Operating an engine with net LCG under wide engine load is very difficult due to low-calorific value of LCG. Nevertheless, LCG can be utilized with gasoline and diesel fuels in dual-fueled SI and CI engines [7,8].

Direct-injection (DI) diesel engines with high thermal efficiency and low CO_2 emissions are proper candidates for utilization of LCG [4,9]. Diesel engines are widely employed in energy generation and cogeneration systems. In DI diesel engines, utilizing LCG would give several effects on engine performance and emissions. At first, existence of hydrogen and carbon monoxide with high heating value and wide combustibility range would lower consumption of diesel fuel [8,10]. The addition of LCG to the intake gas of a diesel engine would also lower the NO_x emission. This is due to a decrease in peak combustion temperature and reduction of oxygen concentration in the intake gas [11,12]. Nevertheless, a decrease in oxygen concentration would cause an increase in smoke, total hydrocarbon (THC), and CO emissions. These influences are observed when a high exhaust gas recirculation (EGR) rate is used in a diesel engine [13]. At the same time, existence of hydrogen in LCG would compensate these effects. Hydrogen with its wide combustibility range is used to activate combustion of other fuels in internal combustion engines. For example, induction of a small amount of hydrogen to a spark-ignited natural gas engine expands the lean limit and improves the thermal efficiency [14]. Moreover, addition of hydrogen obtained from gasoline reforming decreases the combustion period and prevents knocking in a SI engine [15]. Carbon monoxide also enhances the burning rate of hydrocarbon fuels [10]. In diesel engines, hydrogen induction to the intake gas lowers the emissions of greenhouse gases, NO_x and smoke [16–19].

Because of its importance, utilization of LCG in diesel engines has been carried out by some researchers [8]. Their results indicate some of the effects mentioned above. Nevertheless, these studies are carried out using LCG with limited chemical composition. Composition of LCG strongly depends on the type of waste and gasification process [6].

In this study, effects of the LCG amount and composition on diesel-engine performance and emission were parametrically studied using a single-cylinder naturally aspirated DI diesel engine. The LCG was synthesized using a nitrogen and hydrogen mixture, which was introduced into the intake gas. Because of the safety factor, the composition of the LCG was limited to the hydrogen and nitrogen mixture, and CO gas was not used. However, based on the results, possible effects of CO gas in LCG on engine performance were discussed. Experimental results indicate that optimal usage of LCG can reduce diesel fuel consumption and improve emissions.

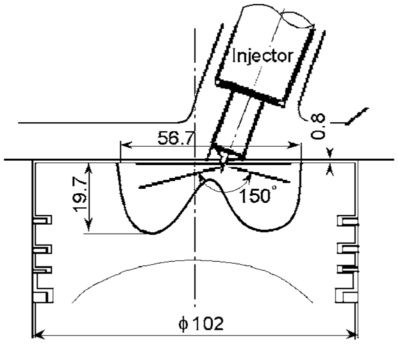
Experimental Apparatus and Procedure

The test engine used was a four-stroke single-cylinder naturally aspirated DI diesel engine (Yanmar NFD-170) with a bore of 102 mm and a stroke of 105 mm. Figure 1 shows the configuration of the cylinder head and combustion chamber. The specification of the injection nozzle is spray angle of 150 deg and four holes with 0.29 mm hole diameter. Figure 2 shows the schematic of the experimental setup. Combustion analysis was carried out measuring in-cylinder pressure every 1 deg CA (crank angle) by a piezoelectric pressure transducer (Kistler 6052A). The diesel fuel used was JIS#2 gas-oil (density of 828 $\mathrm{kg/m^3}$, lower heating value of 44,200 kJ/kg, and cetane number of 55). All experiments were conducted at thermally steady states of the engine at: an inlet cooling water temperature of 80°C, a lubricating oil temperature of 90°C, and an injection timing of $\theta_j = 12$ deg BTDC. Engine speed was kept constant at 1800 rpm.

Effects of low-calorific gases on engine performance were investigated by adding the hydrogen and nitrogen mixture into the engine intake gas. Nitrogen from a high-pressure vessel was introduced into the intake gas using a gas mixer installed at downstream of the surge tank, and hydrogen gas was induced using an orifice nozzle with diameter of 6 mm. Flow rates of both gases

---

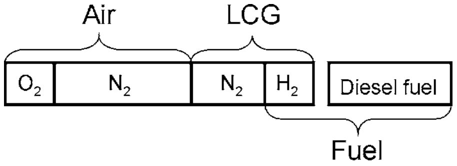
Contributed by the International Gas Turbine Institute of ASME for publication in the Journal of Engineering for Gas Turbines and Power. Manuscript received September 14, 2004; final manuscript received November 2, 2005. Review conducted by D. N. Assanis.



Engine type	DI diesel engine Single cylinder, Water cooled
Borex Stroke	102mm× 105mm
Displacement	857 cc
Compression ratio	17.8:1
Swirl ratio	2.6
Combust. chamber	Toroidal

Fig. 1 Engine specification and combustion chamber

were accurately measured using thermal mass flowmeters. In the experiment, first, the flow rate and composition of the LCG were adjusted and then the amount of diesel fuel injected was increased to achieve the intended engine power. Several gas analyzers were used to clarify the effects of the LCG addition on engine emissions. NO_x emission was measured using a chemiluminescent analyzer (Yanaco ECL-30). Smoke density was measured using a Bosch smoke meter (Zexel DSM-10). Unburned hydrocarbons were measured by a heated flame ionization detector (FID) (Horiba Mexa-1160TFI-H). A gas chromatograph (Yanaco G6800) with thermal conductivity detector was used to measure the amount of unburned hydrogen in engine exhaust. Emissions of CO and CO_2 were measured with a non-dispersive infrared absorption (NDIR) analyzer (Altas-121).



$$r_{LCG} = \frac{V_{LCG}}{V_{Air} + V_{LCG}} \qquad r_H = \frac{V_{H2}}{V_{LCG}}$$

Fig. 3 Definition of LCG amount and composition

Results and Discussion

Effects of LCG Addition Amount on Engine Performance at High Engine Loads. At first, the effects of the LCG amount and composition on engine performance and emissions were investigated by varying the LCG amount r_{LCG} and hydrogen concentration r_H in the range of 0.1–0.25 and 0–0.3, respectively. As shown in Fig. 3, r_{LCG} corresponds to the volume ratio of low-calorific fuel in the intake gas and r_H indicates the volume ratio of hydrogen in the low calorific gas. In this case, the engine load was kept constant at brake mean effective pressure of $p_e=0.6$ MPa. As shown in Fig. 4, increasing the amount of LCG and hydrogen concentration in the intake gas does not affect the brake thermal efficiency η_e so much. Here, η_e corresponds to the brake engine thermal efficiency considering both diesel and low-calorific gaseous fuels. Even at $r_H=0$ and $r_{LCG}=0.25$, when 25% of the intake gas is replaced with nitrogen, η_e is slightly lower than diesel fuel operation (DF). However, at each r_{LCG}, increasing the hydrogen concentration lowers the consumption of diesel fuel be_{DF}. At $r_{LCG}=0.25$, $r_H=0.3$, be_{DF} is about 140 g/kWh, which corresponds to a 40% savings in consumption of diesel fuel. Here, be_{DF} only corresponds to the brake specific fuel consumption of diesel fuel.

Next, the effects of LCG addition on exhaust emission were investigated. Figure 5 reveals NO_x, smoke, THC, and CO concentrations in engine exhaust at the same experimental condition as Fig. 4. Increasing r_{LCG}, improves NO_x emission substantially

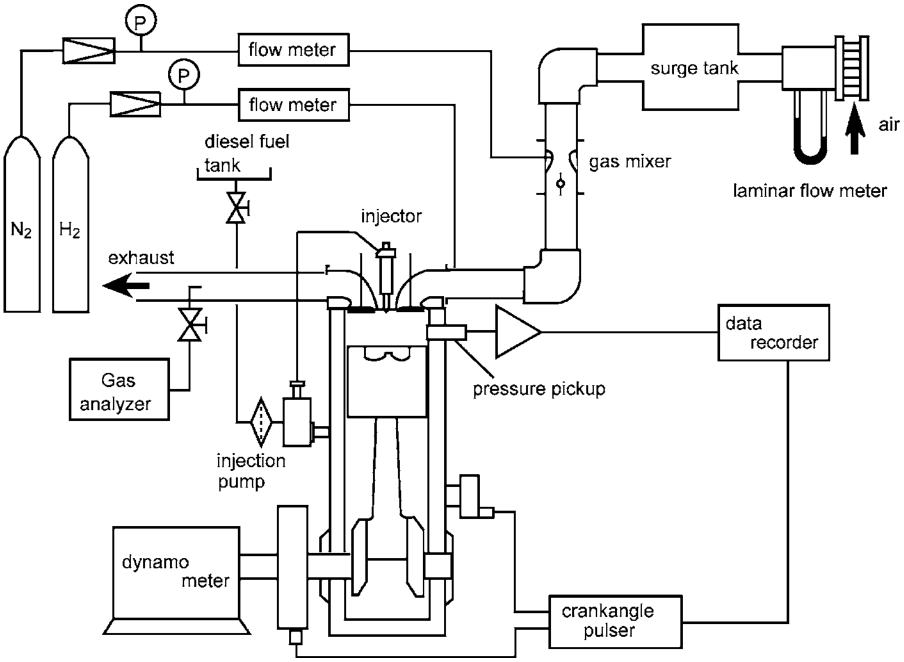


Fig. 2 Experimental setup

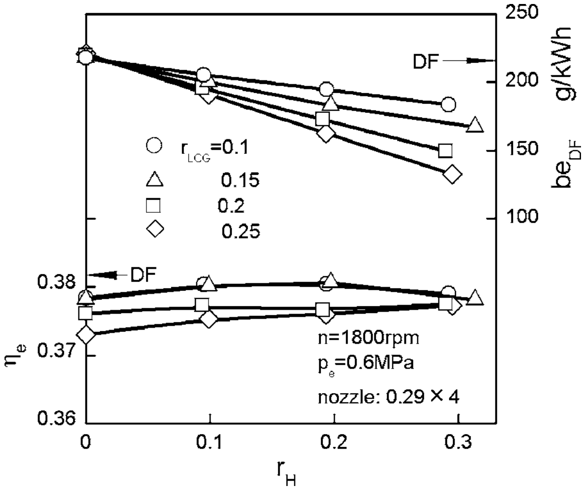


Fig. 4 Effects of LCG addition on engine performance at constant engine load

when compared to that in diesel fuel operation (DF). However, it increases CO, smoke, and THC emissions. Except for THC, by increasing hydrogen concentration to $r_H=0.3$, emissions can be lowered to the same level as the diesel fuel operation (DF).

Figure 6 shows the effects of the LCG addition on in-cylinder pressure p and heat release rate $dq/d\theta$ under the same conditions as Fig. 4. At $r_H=0$ when only nitrogen is added to the engine intake, increasing r_{LCG} prolongs the ignition delay but with minor effects on combustion process. However, at a given r_{LCG}, increasing the hydrogen concentration r_H promotes the premixed and diffusion combustions, giving higher peak combustion pressure. Nevertheless, in this case, little change was observed in the ignition delay. In high LCG addition conditions, a distinct peak in the heat release rate at the diffusion combustion stage can be seen. Increasing r_H increases the peak level and advances its timing. This combustion characteristic would be due to propagation of hydrogen flame initiated by diesel fuel combustion [16–19].

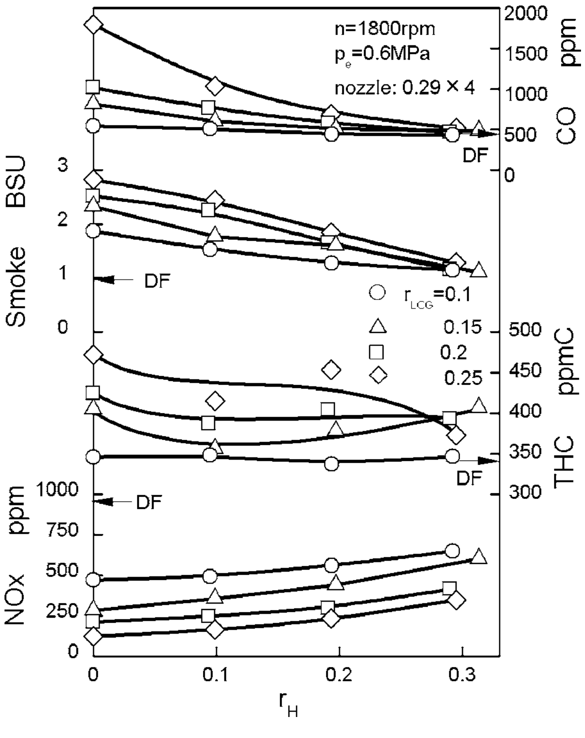


Fig. 5 Effects of LCG addition on exhaust emissions at engine load of p_e=0.6 MPa

From the above results, it can be summarized that the addition of LCG with small hydrogen content, lowers the combustion temperature and oxygen concentration, and therefore offers a low NO_x emission. Nevertheless, reduction in oxygen concentration in the engine intake could be the reason for the increase in smoke, CO, and THC emissions. Increase in smoke could also be due to lower rate of soot oxidation because of a lower combustion temperature [20,21]. Increasing the hydrogen concentration in LCG promotes the premixed and diffusion combustions and improves smoke, CO, and THC emissions.

LCG Addition Under Various Engine Loadings. Effects of the LCG addition on engine performance and emissions were studied at different engine loads when the amount of LCG was kept constant at $r_{LCG}=0.15$ and hydrogen concentration was varied in the range of $r_H=0-0.3$. As shown in Fig. 7, in diesel operation (DF), thermal efficiency η_e increases with p_e and peaks at $p_e=0.6$ MPa. This trend is very similar when LCG gas is used. At $p_e=0.4-0.6$ MPa, variation of hydrogen content does not affect the thermal efficiency η_e. However, at $p_e<0.4$ MPa, increasing hydrogen content lowers the thermal efficiency η_e. However, as shown in the top figure, an increase in hydrogen content still lowers the consumption of diesel fuel be_{DF}.

Next, combustion analysis was carried out to clarify the negative effects of LCG addition on thermal efficiency at low engine loads. Figure 8 shows the effects of hydrogen concentration in LCG on in-cylinder pressure p and rate of heat release $dq/d\theta$ at engine load of $p_e=0.2$ MPa in which reduction in thermal efficiency was observed. This figure also includes the combustion process for $p_e=0.6$ MPa in which thermal efficiency is unchanged by the LCG addition. At $p_e=0.2$ MPa, the addition of LCG delays the ignition and lowers the amount of heat released in premixed combustion giving lower peak combustion pressure. In this case, diffusion combustion is not affected by the LCG addition. However, at high engine load of $p_e=0.6$ MPa, increasing hydrogen in LCG promotes both premixed and diffusion combustions.

Negative effects of the LCG addition on thermal efficiency in low engine load were investigated by measuring the amount of unburned hydrogen in the engine exhaust. Figure 9 indicates the concentration of unburned hydrogen in engine exhaust UH_2 and the combustion efficiency of hydrogen (calculated based on UH_2), against engine load p_e at $r_{LCG}=0.15$ and $r_H=0.3$. It can be seen that at $p_e<0.4$ MPa, the concentration of unburned hydrogen UH_2 in the exhaust gas is relatively high and at $p_e=0.2$ MPa, about 20% of the hydrogen added to the intake gas leaves the engine without combustion.

Figure 10 illustrates the equivalence ratios of hydrogen ϕ_H and diesel fuel ϕ_{DF} against p_e for $r_{LCG}=0.15$ and $r_H=0\sim0.3$. ϕ_H and ϕ_{DF} are determined based on the entire intake oxygen. As shown here, the hydrogen concentration of $r_H=0.1-0.3$ in LCG corresponds to the equivalence ratio of $\phi_H=0.05-0.125$. In low engine loads, combustion of diesel fuel, which is the ignition source for hydrogen premixture, is spatially limited and unable to burn the very lean hydrogen mixture distributed in the whole combustion chamber. This would be the reason for low thermal efficiency of engine at low loads when LCG was used.

Next, the effects of LCG addition on emission were investigated under various engine loads. Figure 11 shows exhaust emissions at the same experimental conditions as Fig. 7. The increase in NO_x concentration with p_e for LCG is similar to that in diesel fuel operations DF. However, its level is much lower. Concentrations of THC, CO, and smoke increase with LCG addition. However, when hydrogen concentration is increased up to $r_H=0.3$, the emission levels reach the diesel fuel operation (DF) level.

Summarizing the above results, it can be suggested that in order to achieve high thermal efficiency and clean engine emission, op-

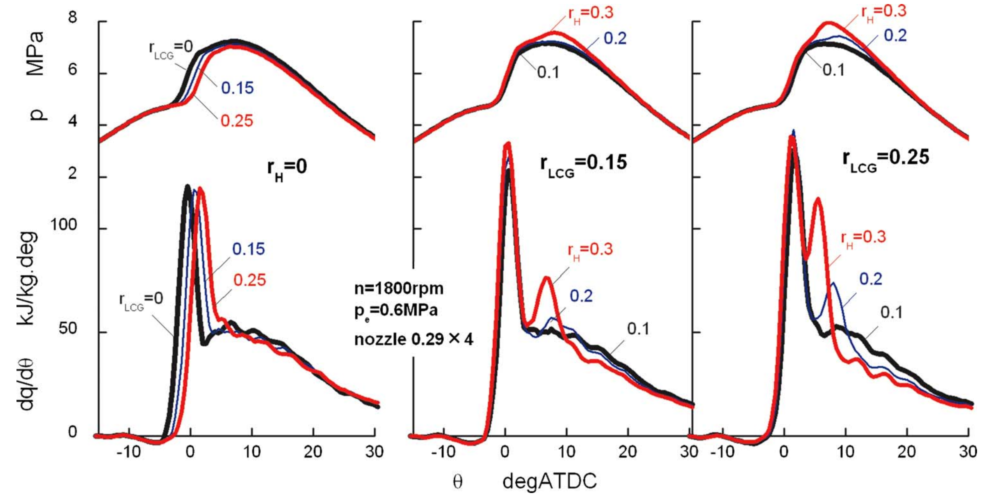


Fig. 6 Effects of LCG addition on in-cylinder pressure *p* and heat release rate $dq/d\theta$ at engine load of $p_e=0.6$ MPa

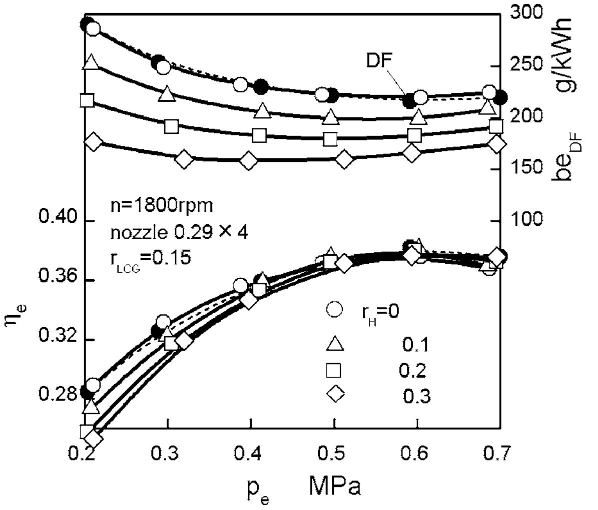


Fig. 7 Effects of LCG addition on thermal efficiency η_e and consumption of diesel fuel be_{DF} under various engine loads

eration of an engine with LCG would be better to be limited to high engine loads. And also, hydrogen concentration of at least 30% is required.

LCG Addition Under Various Injection Timing. In DI diesel engines, injection timing greatly affects the engine performance and emissions. Generally, for diesel fuel operation, retarding injection timing offers lower NO_x emission but higher smoke emissions. Advancing injection timing gives opposite results. An investigation was carried out to see the effects of injection timing when LCG is added to the engine intake gas.

In the test engine used here, variation of the injection timing is very difficult. Therefore, the original injection pump was replaced with a jerk-type pump (Zexel Type A), which offers a high flexibility in the injection timing. In addition, nozzle was changed to a (DLL-P)-type nozzle with four holes and hole size of 0.24 mm to match the small plunger size of the newly used injection pump. Figures 12 and 13 show the engine performance and emissions when injection timing θ_j was varied in the range of −15 to

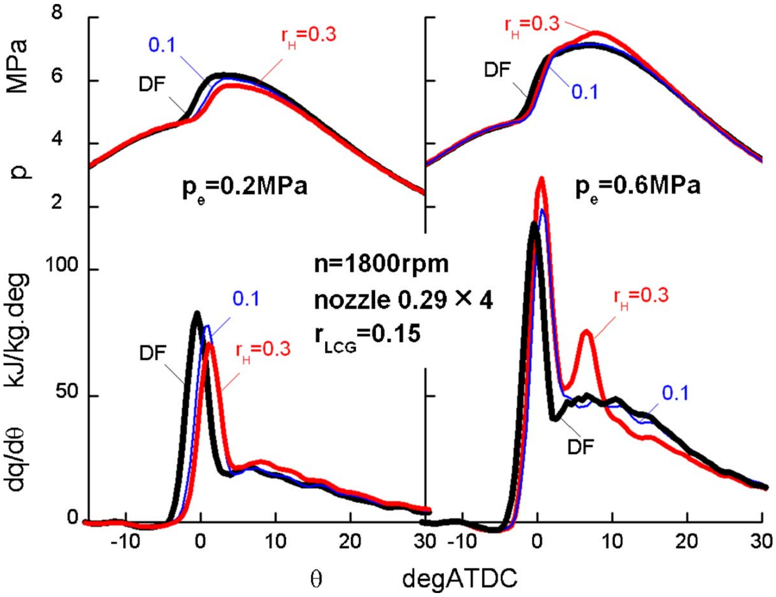


Fig. 8 Effects of LCG addition on in-cylinder pressure *p* and heat release rate $dq/d\theta$ under low and high engine loads

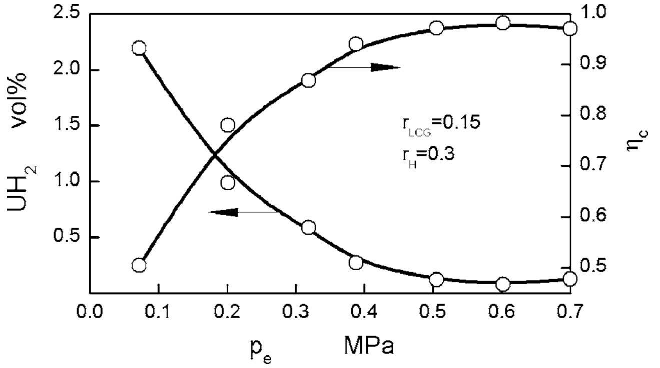


Fig. 9 Combustion efficiency of hydrogen η_c and concentration of unburned hydrogen UH$_2$ in engine exhaust against engine load

−7.5 deg ATDC at engine load of p_e=0.6 MPa. In this case, the amount and composition of the LCG were kept constant at r_{LCG} =0.15 and r_H=0.3; respectively. Results indicate that advancing injection timing improves thermal efficiency η_e. This trend is similar for both fuels. As shown in Fig. 13, advancing the injection timing, improves smoke, THC, and CO emissions. Nevertheless, NO$_x$ emission is worsened. However, for the LCG addition, even under advanced injection conditions NO$_x$ emission level is still much lower than that in diesel fuel operation.

This result reveals that using LCG under advanced injection conditions, one can achieve low NO$_x$ emissions—almost the same thermal efficiency—of smoke, THC, and CO emissions as the diesel fuel operation.

Conclusions

This study demonstrates the nearly optimized usage of low-calorific gaseous fuel in a direct injection diesel engine. The results are summarized as follows:

1. Intake addition of LCG with proper amount and composition does not affect the thermal efficiency at middle and high engine loads. Nevertheless, at low engine loads, thermal efficiency is decreased with the LCG addition. Regardless of thermal efficiency, the LCG addition offers a substantial saving in diesel fuel consumption.
2. At middle and high engine loads, promotion of the premixed and diffusion combustions due to the LCG addition is a reason for the high thermal efficiency. However, at low engine loads, combustion of small amounts of injected diesel fuel is not sufficient to burn a very lean mixture of hydrogen. This would be a reason for the low thermal efficiency of engine when LCG is utilized.
3. The LCG addition improves NO$_x$ emission due to the high amount of nitrogen added into the intake gas. Increasing hydrogen concentration up to 30% lowers the NO$_x$ improvement. However, level of NO$_x$ emission is still lower than that in diesel fuel operation.
4. Increasing the LCG concentration in intake gas gives an increase in smoke, THC, and CO emissions. This would be due to the reduction in oxygen concentration, which promotes soot formation and lowers the soot oxidation rate. Nevertheless, increasing the hydrogen concentration up to 30% compensates for these disadvantages.
5. At high engine loads, advancing the injection timing improves thermal efficiency and smoke, THC, and CO emis-

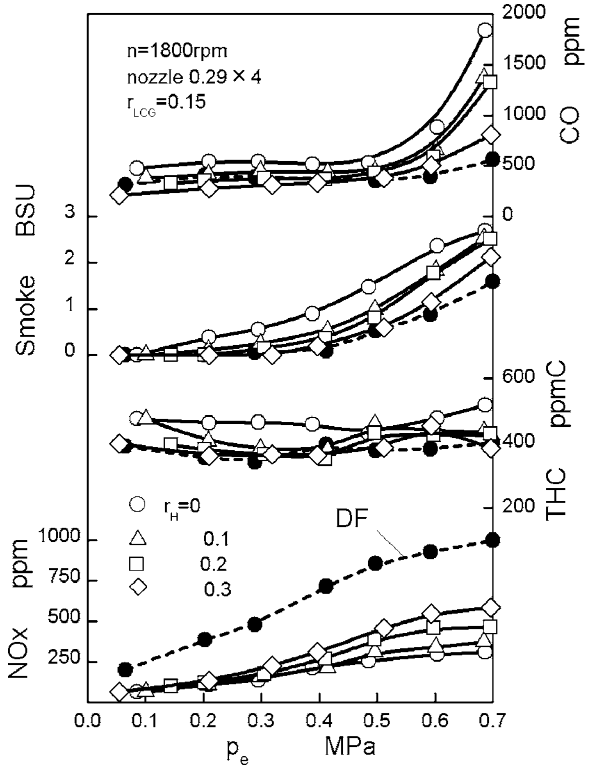


Fig. 11 Effects of LCG composition on exhaust emissions under various engine loads

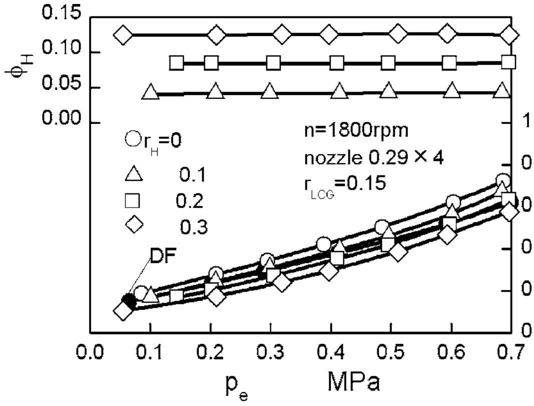


Fig. 10 Equivalence ratios of diesel fuel ϕ_{DF} and hydrogen ϕ_H under various engine loads and LCG compositions

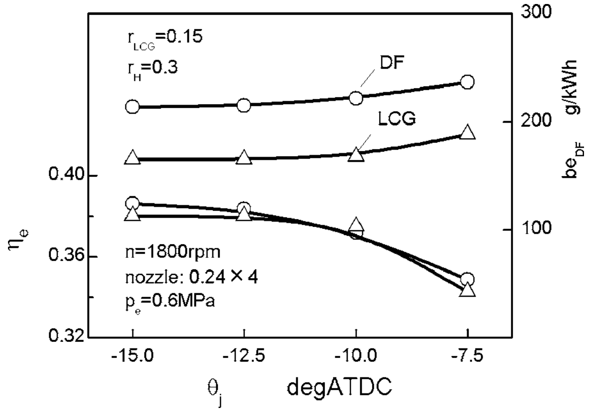


Fig. 12 Effects of injection timing on brake thermal efficiency η_e and diesel fuel consumption be_{DF}

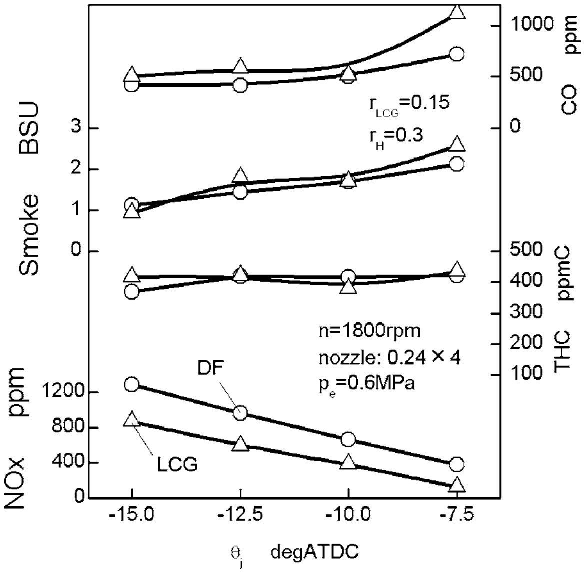


Fig. 13 Effects of injection timing on exhaust emissions

sions for both diesel and LCG fuels. In addition, the LCG addition offers lower NO_x emissions where other engine performance and emissions characteristics are the same as diesel fuel operation.

Finally, it must be mentioned that, in most studies (including this study), the effects of the LCG addition on engine performance and emissions were investigated using diesel engines with basic technologies. Nowadays, diesel engines employed in power generation are equipped with recent combustion control technologies, such as optimized combustion chamber geometry, high-pressure fuel injection system, cold EGR, supercharger, etc. The authors believe that the tendency to use low-calorific gas with low NO_x emissions in modern diesel engines, generally, with low smoke, CO, and THC emission would offer even better performance and emission characteristics than what is reported here in this study. These technologies are important also regarding CO gas in low-calorific gas. As mentioned earlier in this paper, due to the safety factor, CO gas was not added to the engine intake. Carbon monoxide, with a reasonable heating value and wide combustibility range, is an important substance in LCG. Higher CO concentrations in LCG would offer higher savings in diesel fuel consumption. However, if its concentration is exceeded too much, then this would reflect on CO emission in engine exhaust. The authors believe that using LCG in diesel engines equipped with recent combustion control technologies would compensate this negative effect.

Acknowledgment

The authors wish to thank Naohiro Mizutani and Yasuyuki Nakai, graduate students of Kyoto University, for valuable laboratory work. This study was partially supported by a Grant-in-Aid for the 21st Century COE program “Establishment of COE on sustainable Energy System” from Japanese Ministry of Education, Culture, Sports, Science and Technology. The authors would like also to thank Yanmar Co. Ltd. for supporting this work.

References

[1] SolidWaste.com, 2004, statistics of world waste productions.

[2] Department of Energy DOE, 2004, International Energy Outlook, www.eia.doe.gov/oiaf/ieo/electricity.html

[3] Björklund, A., Melaina, M., and Keoleian, G., 2001, “Hydrogen as a Transportation Fuel Produced From Thermal Gasification of Municipal Solid Waste: An Examination of Two Integrated Technologies,” Int. J. Hydrogen Energy, **26**, pp. 1209–1221.

[4] Bridgwater, A. V., Toft, A. J., and Brammer, J. G., 2002, “”A Techno-Economic Comparison of Power Production by Biomass Fast Pyrolysis With Gasification and Combustion, Renewable Sustainable Energy Rev., **6**, pp. 181–248.

[5] Pian, C., and Yoshikawa, K., 2001, “Development of a High-Temperature Air-Blown Gasification System,” Biosources Tech., pp. 231–241.

[6] Yamashita, H., Yoshikawa, K., Hashimoto, Y., and Matsuoka, K., 2003, “Development of Micro-Gasification and Power Generation System for On-Site Treatment of Solid Waste,” *Proc. of Int. Conference on Power Engineering-03, ICOP-03*, JSME-KSME, Kobe, pp. 3-335–3-339.

[7] Shudo, T., Nagano, T., and Kobayashi, M., 2001, “Combustion Characteristics of Pyrolysis Gas of Automobile Shredder Dust in an IC Engine,” European Automotive Congress, SAITS01002, Bratislava, Slovakia.

[8] Tabata, D., Yoshikawa, K., and Min, T., 2003, “Performance of a Dual-Fueled Diesel Engine Driven by Low Calorific Gas,” *Proc. Int. Conf. on Power Engineering, ICOPE-03*, JSME-KSME, Kobe, pp. 1-65–1-68.

[9] Heywood, J. B., 1988, *Internal Combustion Engine Fundamentals*, McGraw-Hill, New York.

[10] Sung, C. J., and Huang, Y., 2001, “Effects of Reform Gas Addition on the Laminar Flame Speeds and Flammability Limits of n-Butane and iso-Butane Flames,” Combust. Flame, **126**, pp. 1699–1713.

[11] Yu, R. C., and Shahed, S. M., 1981, “Effects of Injection Timing and Exhaust Gas Recirculation on Emissions From a D. I. Diesel Engine,” Trans. ASAE, **90**, 811234.

[12] Satoh, K., Zhang, L., Hatanaka, H., Takatsuki, T., and Yokota, K., 1997, “Relation Between NOx and SM Emissions From DI Diesel engine With EGR,” JSAE Rev. **18**, pp. 369–375.

[13] Mohammadi, A., Ishiyama, T., Kawanabe, H., and Horibe, N., 2004, “An Optimal Usage of Recent Combustion Control Technologies for DI Diesel Engine Operating on Ethanol-Blended Fuels,” SAE paper No. 2004-01-1866.

[14] Shioji, M., Eguchi, S., Kitazaki, M., and Mohammadi, A., 2004, “Knock Characteristics and Performance in an SI Engine With Hydrogen and Natural-Gas Blended Fuels,” SAE Paper No. 2004-01-1929.

[15] Shinagawa, T., Okumura, T., Furuno, S., and Kim, K. O., 2004, “Effects of Hydrogen Addition to SI Engine on Knock Behavior,” SAE Paper No. 2004-01-1851.

[16] Gopal, G., Srinivasa Rao, P., Gopalakrishnan, K. V., and Murthy, B. S., 1982, “Use of Hydrogen in Duel-Fuel Engines,” Int. J. Hydrogen Energy, **7**(3), pp. 267–272.

[17] Varde, K. S., and Frame, G. A., 1983, “Hydrogen Aspiration in a Direct Injection Type Diesel Engine-Its Effects on Smoke and Other Engine Parameters,” Int. J. Hydrogen Energy, **8**(9), pp. 549–555.

[18] Lambe, S. M., and Watson, H. C., 1993, “Optimizing the Design of a Hydrogen Engine With Pilot Fuel Injection,” Int. J. Veh. Des., **14**(4), pp. 370–389.

[19] Tomita, E., Kawahara, N., Piao, Z., Fujita, S., and Hamamoto, Y., 2001, “Hydrogen Combustion and Exhaust Emissions Ignited With Diesel Oil in a Dual-Fuel Engine,” SAE Paper No. 2001-01-3503.

[20] Hentschel, W., and Richter, J. U., 1995, “Time Resolved Analysis of Soot Formation and Oxidation in Direct-Injection Diesel Engine for Different EGR-Rates by an Extinction Method,” SAE Paper No. 952517.

[21] Dec, J. E., and Kelly-Zion, P. L., 2000, “The Effects of Injection Timing and Diluent Addition on Late-Combustion Soot Burnout in a Diesel Engine Based on Simultaneous 2-D Imaging of OH and Soot,” SAE Paper No. 2000-01-0238.

Experimental Investigation of the Effect of Exhaust Gas Recirculation on Lubricating Oil Degradation and Wear of a Compression Ignition Engine

Shrawan Kumar Singh
Environmental Engineering
and Management Program,
Indian Institute of Technology,
Kanpur, Kanpur 208016, India

Avinash Kumar Agarwal

Dhananjay Kumar Srivastava

Department of Mechanical Engineering,
Indian Institute of Technology,
Kanpur, Kanpur 208016, India

Mukesh Sharma
Environmental Engineering
and Management Program,
Indian Institute of Technology,
Kanpur, Kanpur 208016, India

This experimental investigation was aimed to investigate the effect of exhaust gas recirculation (EGR) on wear of in-cylinder engine parts. EGR setup was prepared for a two-cylinder, air-cooled, constant-speed direct-injection compression-ignition engine. Test setup was run for 96 hr *under predetermined loading cycles in two phases; normally, operating condition (i.e., without EGR) and with a fixed EGR rate of 25%. Addition of metallic wear debris in the lubricating oil samples drawn after regular interval from both engine operating phases was investigated. Relatively higher concentrations of all wear metals were found in the lubricating oil of the EGR-operated engine, which indicates higher wear of various engine parts. Weight loss of piston rings used in both phases was compared to quantify the amount of wear of piston rings. To quantify the amount of cylinder wear surface roughness parameters of cylinder liners were measured at three positions (top dead center, mid-stroke, and bottom dead center) on thrust and anti-thrust side. A qualitative analysis was also carried out by taking surface profiles and Scanning Electron Micrographs at same locations.* [DOI: 10.1115/1.2136368]

Introduction

High compression ratio and lean fuel-air mixture give high thermal efficiency to compression-ignition (CI) engines. High fuel economy with high power makes these CI engines popular worldwide. In last decade, diesel engine technology has advanced significantly in terms of speed, power, and efficiency, due to which these engines are gaining popularity even in the passenger car segment [1]. Stringent emission norms for limiting emission of oxides of nitrogen (NO_x) and particulate matter (PM) from diesel engines have been adopted worldwide. In order to meet these norms, different pre and post combustion techniques have been adopted. Exhaust gas recirculation (EGR) is one of the techniques, which is being effectively used to control the NO_x emissions from light-duty diesel engines. Use of EGR in medium and heavy duty diesel engines has been limited so far because of increase in soot, engine wear and oil degradation [2,3]. With the use of EGR, approximately 40–50% reduction in NO_x emission can be achieved [4]. But there is trade-off between NO_x and soot emission. It is believed that as the EGR rate increases, the level of NO_x goes down, but soot increases [5]. The soot generated in the engine can either come out to the atmosphere with the exhaust or may stick to various in-cylinder engine parts or can get adsorbed to the lubricating oil film present on the cylinder walls and finally reach oil sump. The soot can also reach oil sump with blow-by gases. Soot in the lubricating oil accelerates its degradation [6]. Soot present in the lubricating oil consists of up to 90% carbon with traces of metallic elements [7].

In the EGR system, the majority of soot comes to the lubricating oil when it is scrapped down due to motion of rings or blow-by gases [8]. Cadman and Johnson [9] found ten times higher particulate concentration in the lubricating oil drawn from engines using 15% EGR compared to baseline.

Higher soot/carbon generated by the EGR system is the main cause of higher engine wear [10]. Soot particles act as abrasives and are adsorbed by the oil film present on the cylinder liner surface. When these soot particles are deposited at the interface between the piston rings and cylinder liner, a three-body wear mechanism occurs. The three-body wear mechanism involves two surfaces and entrapped particles, and wear occurs at the particle-surface interface. The diameter of soot particles generally ranges from 0.03–0.10 μm, whereas the thickness of boundary layer lubricating oil film is typically <0.025 μm [11,12]. Ishiki et al. [13] confirmed higher wear in piston rings of EGR-operated engines due to accelerated abrasive wear, experimentally.

Another possible reason for higher wear of various parts in the engine operated with EGR is corrosion, which occurs due to the formation of sulfuric acid by reaction of SO_x (formed during combustion) with the condensed water on cylinder surfaces [14]. When the sulfuric acid reaches oil sump, it reduces the total base number (TBN) of lubricating oil thereby affecting properties of the lubricating oil. The carbon (soot) in the lubricating oil reduces the effectiveness of antiwear film. Soot could hinder the additives from reaching the contact area by adsorbing them, thus, reducing the functionality of these additives [15,16].

The cylinder liner of a normally operating engine (operating without EGR) undergoes maximum wear at top dead center (TDC) and bottom dead center (BDC) because of boundary lubrication conditions prevailing at these locations. At these locations, the hydrodynamic lubricating oil film collapses due to near-zero relative velocity between the piston ring and liner [17]. The present experimental study is aimed at quantifying the wear of piston rings and cylinder liners because of the use of EGR in a typical constant-speed diesel engine and comparing it to the same

---

Contributed by the Internal Combustion Engine Division of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received January 14, 2005; final manuscript received September 26, 2005. Review conducted by D. Assanis.

Table 1 Technical specification of the test engines

Manufacturer/Model	Indian Ntl. Diesel Engine Co., (Under License from Hawker Siddley Ind. Ltd., UK) Petter Design/Indec PH2
Engine type	Two-cylinder, four-stroke, air-cooled, direct-injection CI engine
Bore/stroke (mm)	87.3/110
Rated power	9.32 kW @ 1500 rpm
Compression ratio	16.5:1
Displacement volume (cc)	1318
Fuel-injection pressure (MPa)	21

Table 2 Engine test cycle for endurance

Load (%)	Duration (minutes)
No load	20
100	30
50	120
No load	20
75	60
No load	20
100	30
75	60

engine operating with EGR under similar operating conditions. The other parameters, such as lubricating oil, fuel, etc., are kept identical for the entire experiments.

Experimental Setup and Methodology

The experiments were conducted on a constant-speed, two-cylinder, four-stroke diesel engine. This type of engine is typically used for captive power generation and for agricultural irrigation purposes. The technical specifications of the engine are given in Table 1. The exhaust gas recirculation setup was installed on the engine. Hot exhaust gases were recirculated and mixed with fresh air in intake manifold. The schematic diagram of the experimental setup is shown in Fig. 1.

The experiments were conducted in the following two phases:

1. baseline engine operation, i.e., engine operated without EGR under normal operating condition
2. engine operation with a fixed EGR rate of 25%

The engine test runs were conducted for 96 hr in each phase of the experiments. A typical load cycle of 6 hr (continuous) was selected for both phases. The engine test cycle for endurance is shown in Table 2.

A new set of cylinder liners, pistons, and piston rings were installed in the engine in the beginning of each phase of the experiments. SAE 20W/40 lubricating oil was filled in the engine at the start of test run, and no makeup oil was added. In both the phases of the engine experiments, no major breakdown was observed.

Wear of the engine is a slow and continuous process. In the engine, metallic wear debris is washed away by lubricating oil and they accumulate in the oil sump. Hence, wear-particle analysis is a powerful technique to examine the condition of vital engine parts.

Lubricating oil samples were drawn from the engine after every 24 hr of engine run for wear-metal analysis. Lubricating oil samples were analyzed on a flame atomic absorption spectroscopy unit (Make: Varian AAS spectra Model: AA 220FS) for evaluating the metallic composition of wear debris. After completion of both phases of the experiments, the piston rings were weighed in order to evaluate their wear. The liner surface profile was evaluated for various surface characteristics before and after the completion of each phase. A surface profilometer (Make: Mitutoyo, Japan, Model: SJ-301) was used for surface evaluation. Scanning electron microscopy (Make: FEI, Holland, Model: Quanta 200) was done for liner surfaces at TDC, BDC, and midstroke positions for both phases.

Experimatal Results and Discussions

Wear-Metal Analysis of Lubricating Oil Samples. Several sliding and rotating metallic components are involved in engine operation and wear debris originates from these parts. This debris

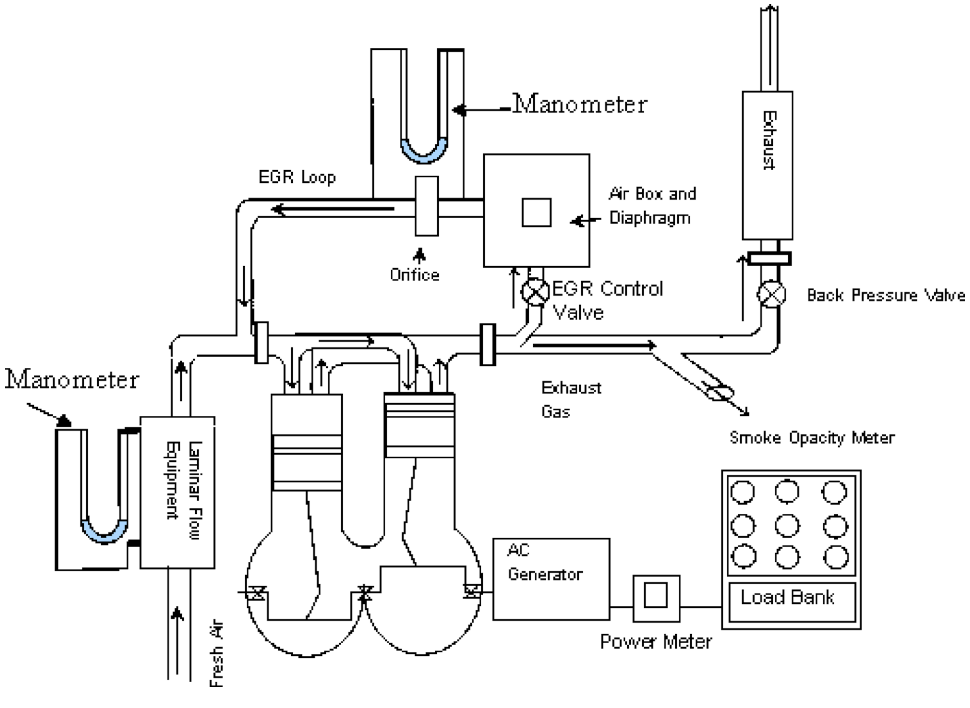


Fig. 1 Schematic diagram of experimental setup

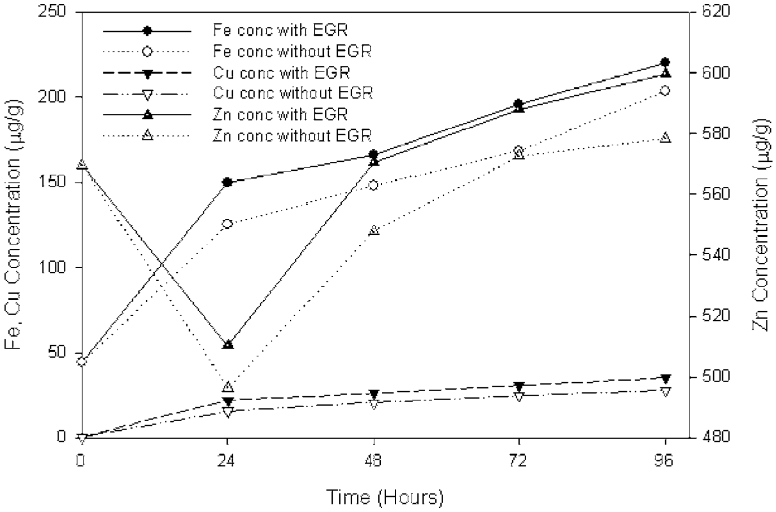


Fig. 2 Fe, Cu, and Cr concentration as a function of lubricating oil usage

finally accumulates in the lubricating oil sump. The oil samples drawn at regular intervals were analyzed for various metals present in the debris, such as Fe, Cu, Cr, Ni, Al, Mn, Pb, Mg, and Zn. For extracting the metals from lubricating oil samples, dry-ashing technique was used [18]. An approximately 5 gm lubricating oil sample was taken in a silica crucible and kept on hot plate at 120°C until it became free from moisture. Then the sample was kept in a muffle furnace at 450°C for 4 hr, thereafter at 650°C for 2 hr. The dried ash was then dissolved in 1.5 mL concentrated HCl, and the mixture was diluted to 100 mL using distilled water. Thus, all the nonorganic components present in lubricating oil gets transferred into the aqueous solution. The samples were then analyzed by atomic absorption spectroscope (AAS) for different metals. The concentration of various wear metal present in lubricating oil samples are shown in Figs. 2–4.

Figures 2–4 show an increasing trend of metallic concentration in the lubricating oil with usage for both experimental phases. The rate of rise in concentrations is initially higher, and then it slows down. The reason for this behavior may possibly be due to the use of new in-cylinder engine parts for each phase. Higher wear of these parts is observed possibly due to higher wear during initial running-in.

Each of these metals can be traced back to several engine parts. Typical sources of these metals present in the lubricating oil are the piston rings, piston, cylinder liner, bearings, connecting rod, crank shaft, valves, lubricating oil additives, etc. Table 3 represents the origin of each metal to different engine parts.

After the initial experimental phase, the rate of metal concentration increase in lubricating oil decreases. The results obtained prove the statement given by Sachs [21] that an engine component has three phases during its life span. The highest possibility of wear of a component is in its initial stage. After few hours of engine operation, the failure possibility reduces and the wear rate decreases. The wear rate again rises when the life of the engine component comes to an end in the last phase, enhancing the possibility of wear leading to failure.

The metallic concentration in the lubricating oil of an engine operated with EGR is always higher than an engine operated without EGR. This indicates higher wear of engines operated without EGR in comparison to engines operates without EGR. Figure 4 shows the variation in Zn concentration as a function of lubricating oil usage. It has been found that the concentration of Zn initially decreases and then increases for both phases. The possible reason for the initial decrease in Zn concentration may be the thermal stressing of the Zn containing additives followed by their evaporation. In the later part of the engine experiments, wear of

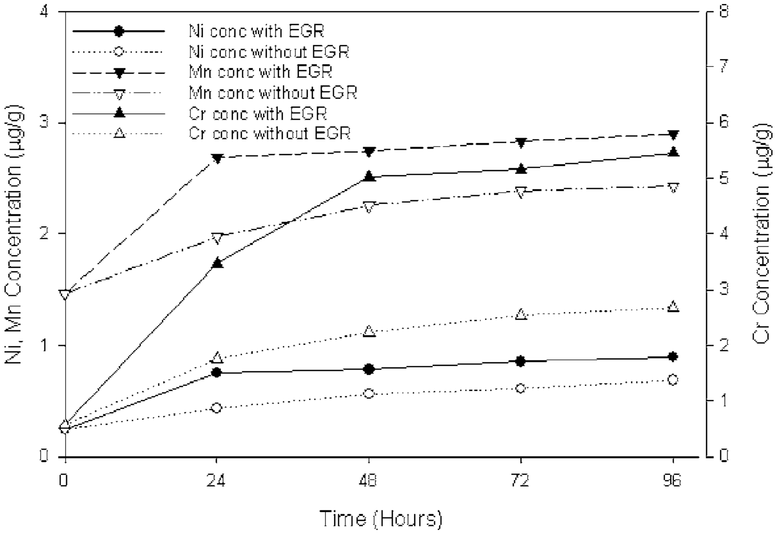


Fig. 3 Ni, Al, and Mn concentration as a function of lubricating oil usage

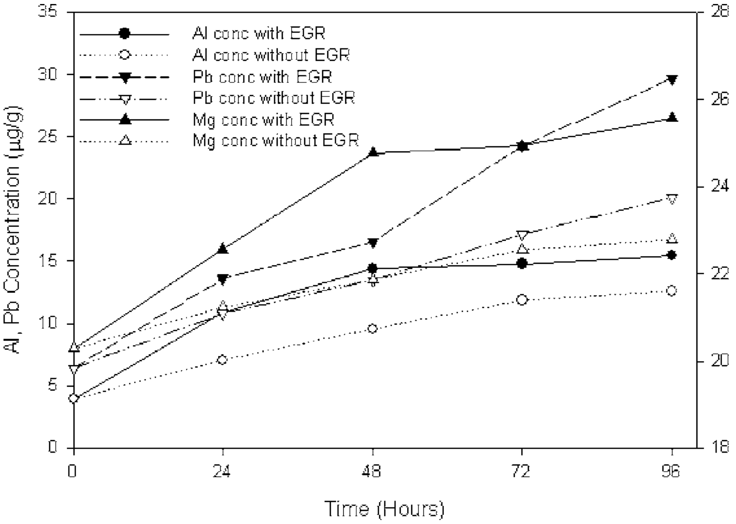


Fig. 4 Pb, Mg, and Zn concentration as a function of lubricating oil usage

various engine parts, such as bearings and galvanized piping, etc., increases Zn concentration in the lubricating oil. Thus, wear metal analysis gives an indication toward higher wear in EGR-operated engines compared to normally operated engines.

Soot Loading of Lubricating Oil. Major part of the soot present in the lubricating oil is in organic carbon form. The lubricating oils from the engine contain negligible amounts of inorganic carbon. The base stock of lubricating oils is organic. Additives are also complex organic metallic compounds. There is very little chance of a reaction of lubricating oil in order to form inorganic carbon while the engine is running. Keeping these points in mind, the lubricating oil samples were analyzed for total carbon (TC) to quantify the addition of soot in the lubricating oil. The change in the amount of carbon present in the used oil compared to fresh oil is assumed to be a reasonable approximation of soot loading of lubricating oil. The change in carbon concentration in the lubricating oil with its usage is shown in Fig. 5. The initial carbon concentration in the lubricating oil was about 88%.

From Fig. 5, it is clear that carbon concentration increases at a faster rate in lubricating oil samples drawn from an EGR-operated engine compared to a normally operated engine. This higher increase in levels of carbon (in other words, soot) in the lubricating oil drawn from an EGR-operated engine may reduce the lubricating efficiency of the oil, resulting in increased wear of vital engine parts and a reduced useful life of lubricating oil.

Wear of Piston Rings. In both phases of the experiments, new engine parts, including piston rings, were installed before starting the test. To quantify the amount of wear, piston rings were weighed before engine assembly and after 96 hr of engine run. The percentage weight loss of rings for both phases of experiments are shown in Fig. 6.

It has been observed that the top compression ring (CR1) of the engine operating without EGR has maximum weight loss among three compression rings (0.50% of its initial weight). The top compression ring faces the highest thrust of combustion gases and

Table 3 Typical sources of wear metals in lubricating oil [19,20]

Elements	Sources
Aluminum	Piston, bearing, dirt
Chromium	Compression rings, coolant, crank shaft, bearings, plating of cylinder liners
Copper	Bearing, piston rings
Iron	Wear from engine block, cylinder liner, rings, crankshaft, anti fiction bearings
Lead	Bearings
Magnesium	Bearings, cylinder liner
Nickel	Piston rings, valves
Zinc	Bearings, platting, brass components
Manganese	Steel shafts, valves

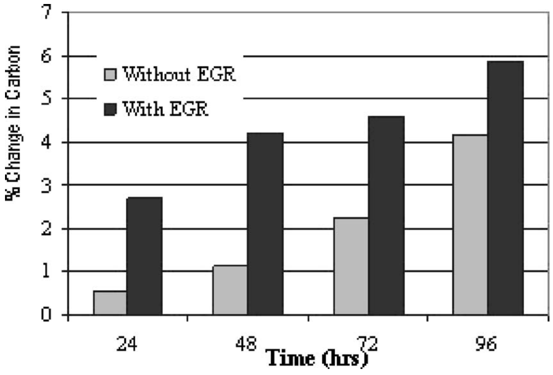


Fig. 5 Percent change in carbon content as function of lubricating oil usage

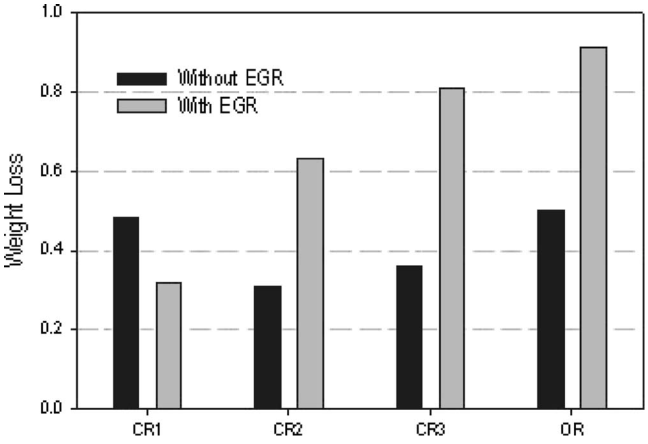


Fig. 6 Percent weight loss of piston rings after 96 hr of engine run

Table 4 Roughness parameters of engine operating with and without EGR

Parameter (μm)	Initial	After 96 hr of engine run: TDC	Midstroke	BDC
		Thrust side		
R_a	0.90 (1.18)[a]	0.75 (0.66)	0.50 (0.34)	0.60 (0.20)
R_q	1.15 (1.57)	1.03 (0.87)	0.78 (0.59)	0.64 (0.37)
R_t	7.18 (12.00)	6.49 (5.31)	6.98 (4.57)	7.37 (3.45)
R_p	2.78 (3.57)	0.98 (0.18)	1.81 (0.67)	1.24 (0.37)
R_v	3.01 (8.25)	5.51 (3.51)	5.17 (3.90)	6.13 (3.08)
		Anti-thrust side		
R_a	0.81 (1.72)	0.49 (0.47)	0.65 (0.60)	0.58 (0.16)
R_q	1.05 (2.59)	0.99 (0.65)	0.76 (0.85)	0.77 (0.34)
R_t	7.84 (21.98)	7.57 (5.27)	6.67 (5.43)	4.87 (3.53)
R_p	3.11 (10.76)	1.35 (2.20)	1.77 (1.23)	1.02 (0.36)
R_v	4.51 (11.22)	6.22 (3.07)	4.90 (4.20)	3.85 (3.17)

[a] The values in the parentheses are for EGR phase experiment.

works in the highest temperature zone; therefore, the top compression ring undergoes maximum wear. The weight loss of oil ring was found comparable to the top compression ring (approximately 0.50%). In the engine using EGR, the top compression ring faces the lowest weight loss compared to other rings (~0.30%), and the oil ring (OR) faces highest weight loss (~0.90% of initial weight). A possible reason for this may be a lower temperature of the combustion chamber of the engine using EGR, which exposes the top compression ring to lower gas thrust and temperature, causing lower wear. However, the wear of the second and third compression rings and the oil ring were comparatively higher for engines using EGR. The possible reason for this may be the presence of a higher amount of soot and wear debris at the ring-liner interface in the engine using EGR. The oil film present in the ring-liner interface region of second and third rings is significantly thicker compared to that of top ring-liner interface region.

Wear of Cylinder Liner Surfaces. In both sets of experiments, new cylinder liners were installed in the engine before starting the experiments. To check the wear of cylinder liner material during the engine run with EGR and without EGR, surface profiles were taken before and after 96 hr of the engine test run, at identical locations. The surface profiles were evaluated at three locations namely top dead center (TDC), midstroke, and bottom dead center (BDC) on thrust and anti-thrust side. TDC and BDC are the most important locations of the cylinder liner with possibility of high wear. The surface profilometer evaluates the surface textures and gives a number of surface roughness parameters, such as R_a, R_q, R_v, R_p, R_t, etc. The evaluation length was 12.5 mm. Profile magnification was 10× in the horizontal direction and 2000× in the vertical direction. The average roughness R_a is the area between the roughness profile and its mean line, or the integral of the absolute value of the roughness profile height over the evaluation length. R_a does not give a clear picture about a surface. Surfaces may have the same value of R_a, but they may be quite different in the shape of the profile. Even if two profiles have similar shapes, they may have different spacing between features. Hence, it is needed to distinguish between surfaces that differ in shape or spacing. Other parameters for a surface that measures peaks, valleys, profile shapes and spacing, etc., need to be calculated. Root-mean-square roughness (R_q) of a surface is calculated from another integral of the roughness profile. The peak roughness R_p is the height of the highest peak in the roughness profile over the evaluation length. Similarly, R_v is the depth of the deepest valley in the roughness profile over the evaluation length. The total roughness R_t is the sum of these two, or the vertical distance from the deepest valley to the highest peak. In practice as R_a is a measure of the departure of the profile from the mean line and R_q is the rms roughness, they provide similar information about a surface. In practice, R_q is a more statistically significant parameter, whereas R_a has the advantage of being more a commonly used parameter.

The roughness parameters of cylinder liner for engines operated with and without EGR on thrust and anti-thrust sides are shown in Table 4, and surface profiles are shown in Figs. 7 and 8. It can be

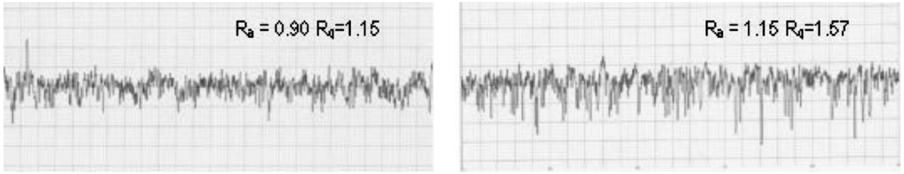


Fig. 7 Surface profile of fresh liner (*a*) without EGR (*b*) with EGR

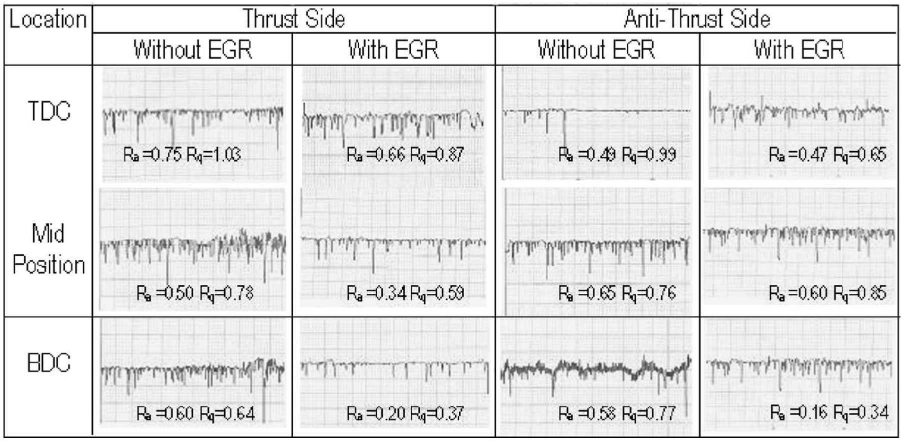


Fig. 8 Surface profiles of cylinder liner surfaces of engine operating with and without EGR after 96 hr

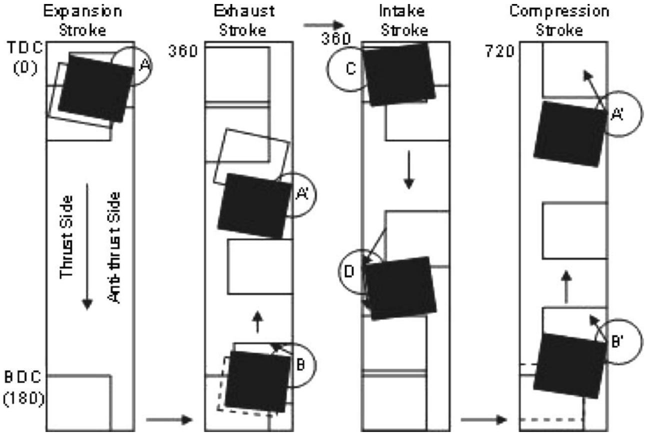


Fig. 9 Representation of piston motion [22]

observed that the wear of the liner at TDC is consistently more than at BDC and midstroke positions when engines operate without EGR. The lower values in Table 4 indicate a flatter and smoother surface, reflecting higher wear of the initial surface. The wear at TDC is more because this zone of cylinder liner faces the highest temperature due to hot, high-pressure combustion gases. TDC faces the highest load and relatively low piston speeds so boundary lubrication at TDC location breaks down, and, possibly metal-to-metal contact takes place, enhancing the wear.

It can also be observed that wear of the cylinder liner is higher on the anti-thrust side compared to the thrust side. This may happen because of the piston tilt during its strokes. In the four strokes of engine operation, the piston touches the thrust side of cylinder liner surface during intake stroke, when temperature and pressure of engine are low. For the remaining three strokes, the piston touches the anti-thrust side of cylinder liner surface, as shown in Fig. 9.

It can be observed from Fig. 8 that the BDC position of the cylinder liner faces higher wear compared to TDC and midstroke positions, when an engine runs with EGR. In the engine using EGR, the combustion temperatures are relatively lower compared to those without EGR conditions; hence, the engine powerpack gets exposed to less severe conditions causing relatively lower wear at TDC on the thrust side of the liner. While on anti-thrust side, since the piston touches the cylinder liner for three strokes (Fig. 9), inspite of low temperatures inside the combustion chamber, significantly higher wear takes place. At BDC position, boundary lubrication exists and the lubricating oil of an EGR system contains a higher amount of wear debris, which may also be responsible for excessive wear at BDC. The wear of thrust and anti-thrust sides were almost the same in all three locations.

Scanning Electron Microscopy. In failure analysis, scanning electron microscopy (SEM) is a natural extension of optical microscopy. The combination of higher magnification, larger depth of focus, greater resolution, and ease of sample observation makes it one of the most heavily used instruments in research today.

The wear of cylinder liner surfaces was compared by scanning electron microscopy after 96 hr of engine operation. The magnification used is 400×. The SEM micrographs were taken at three locations (TDC, midstroke, and BDC) on both thrust and anti-thrust sides for both phases of the experiment. Figure 10 shows the micrographs of cylinder liner surfaces of both, normally operated engine and EGR-operated engine. It has been observed that in the normally operated engine, high wear was observed at TDC positions on both sides. Midstroke positions show significantly lower wear, and wear on anti-thrust side was higher than on the thrust side.

These observations (Fig. 10) substantiate the results of roughness parameters (Table 4) and surface profiles (Fig. 8). In an EGR-operated engine, the highest wear has taken place at BDC as obtained by surface profiles and also roughness parameters. The cross-hatched honing marks completely disappear at these locations, showing severe damage to the surface.

Conclusions

An experimental investigation was carried out to quantify the wear in an EGR-operated diesel engine vis-à-vis a normally operated diesel engine. Wear metals in lubricating oil, which originate from different moving parts, were analyzed. Higher wear metals were obtained in lubricating oil of an EGR-operated engine, suggesting increased engine wear because of EGR. This is attributed to increased soot contamination of the lubricating oil.

The wear pattern of piston rings and cylinder liner surfaces were also analyzed. It has been observed that piston rings of an EGR-operated engine face a very different trend than normally operated engine. The top compression ring of EGR-operated engine undergoes 0.1% lower wear than a normally operated engine, while the oil ring of an EGR-operated engine faces 0.4% higher wear than its counterpart. The wear pattern of surfaces of cylinder liner was evaluated for surface profile and various roughness parameters on the thrust and anti-thrust sides at three locations. In a

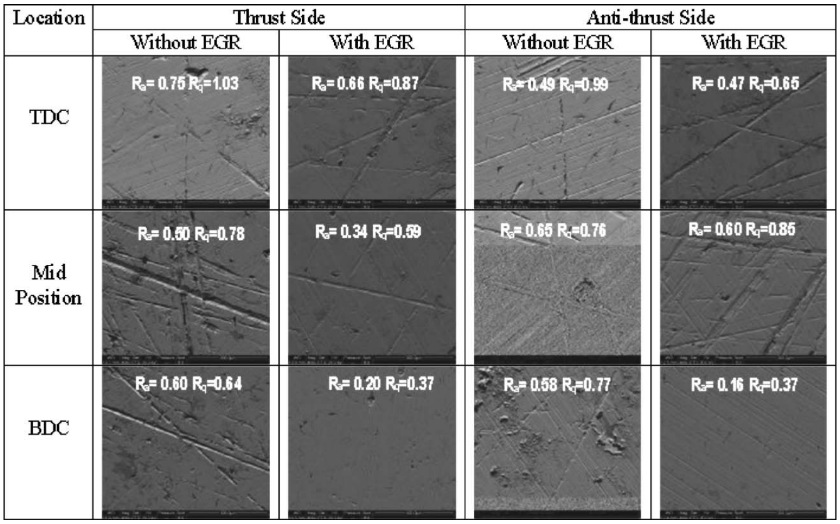


Fig. 10 SEM of cylinder liner surfaces of engine operating with/without EGR

normally operated engine, TDC location faced the highest wear on the anti-thrust side, whereas in an EGR-operated engine, highest wear took place at the BDC position and the wear was almost similar on the thrust and anti-thrust sides. The results obtained were also supported by scanning electron micrographs. The results of the experiments conducted on twin-cylinder contact-speed direct-injection diesel engines using SAE 20W40 lubricating oil and diesel #2 fuel suggest that using EGR to control NO_x enhanced wear of the engine.

Acknowledgment

The authors acknowledge Mohan Lal Saini, Yogesh Pathak, Rais, and A.P. Yadav of Internal Combustion Engine Laboratory, Department of Mechanical Engineering, IIT, Kanpur, for their help in conducting the experiments. Assistance and suggestions of Mritunjay Shukla are appreciated and acknowledged. Research grant from the Ministry of Human Resource Development, Government of India, is acknowledged.

References

[1] Walsh, M. P., 2000, "Global Trends in Diesel Emissions Regulation—A 2001 Update," SAE Paper No. 2000-01-0183.
[2] Needham, J. R., Doyle, D. M., and Nicol, A. J., 1991, "The Low NOx Truck Engine" SAE Paper No. 910731.
[3] Doyle, D., 2002, CTC Analytical Services "EGR System and Lubricating Oil in Diesel Engines," Practicing Oil Anal. Mag., July.
[4] Khair, M., Lemaire, J., and Fischer, S., 2000, "Integration of Exhaust Gas Recirculation, Selective Catalytic Reduction, Diesel Particulate Filters, and Fuel-Borne Catalyst for NOx/PM Reduction," SAE Paper No. 2000-01-1933.
[5] Abd-Alla, G. H., 2002, "Using Exhaust Gas Recirculation in Internal Combustion Engines: A Review," Energy Convers. Manage., **43**, pp. 1027–1042.
[6] Trujillo, G., Noria Latin America, 2004, "Resetting Oil Analysis Parameters for Changing Diesel Engines," Practicing Oil Anal. Mag., Jan.
[7] Clague, A. D. H., Doonet, J. B., Wang, T. K., and Peng, J. C. M., 1999, "A Comparison of Diesel Engine Soot With Carbon Black," Carbon, **37**, pp. 1553–1565.
[8] Tokura, N., Terasaka, K., and Yasuhara, S., 1982, "Process Through Which Soot Intermixes Into Lubricating Oil of a Diesel Engine With Exhaust Gas Recirculation," SAE Paper No. 820082.
[9] Cadman, W., and Johnson, J. H., 1986, "The Study of the Effect of Exhaust Gas Recirculation on Engine Wear in a Heavy Duty Diesel Engine Using Analytical Ferrography," SAE Paper No. 860378.
[10] Dennis, A. J., Garner, C. P., Taylor, D. H. C., 1999, "The Effect of EGR on Diesel Engine Wear," SAE Paper No. 1999-01-0839.
[11] Mainwaring, R., 1997, "Soot and Wear in Heavy Duty Diesel Engines," SAE Paper No. 971631.
[12] Liang, H., Totten, G. E., and Webster, G. M., "Lubrication and Tribology Fundamentals," *Manual 37: Fuels and Lubricants Hand Book*, ASTM, Metals Park, OH, pp. 909–960.
[13] Ishiki, K., Oshida, S., Takiguchi, M., and Urabe, M., 2000, "A Study of Abnormal Wear in Power Cylinder of Diesel Engine With EGR—Wear Mechanism of Soot Contaminated in Lubricating Oil," SAE Paper No. 2000-01-0925.
[14] Akiyama, K., Masunaga, K., Kado, K., and Yoshioka, T., 1987, "Cylinder Wear Mechanism in an EGR—Equipped Diesel Engine and Wear Protection by the Engine Oil," SAE Paper No. 872158.
[15] Rounds, F. G., 1977, "Carbon: Cause of Diesel Engine Wear?" SAE Paper No. 770829.
[16] Rounds, F. G., 1981, "Soot From Used Diesel Engine Oils—Their Effects on Wear as Measured in 4-Ball Wear Tests," SAE Paper No. 810499.
[17] Needalman, W., and Madhavan, P., 1988, "Review of Lubricant Contamination and Diesel Engine Wear," SAE Paper No. 881827.
[18] Palus, J. Z., 1998, "Examination of Used Motor Oils by Flame AAS for Criminalistic Purposes: A Diagnostic Study," Forensic Sci. Int., **91**, pp. 171–179.
[19] Nadkarni, R. K., "Elemental Analysis," *Manual 37: Fuels and Lubricants Hand Book*, ASTM, Metals Park, OH, pp. 707–716.
[20] Agarwal, A. K., 1999 "Performance Evaluation and Tribological Studies on a Biodiesel-Fuelled Compression Ignition Engine," Ph.D. thesis, Centre for Energy Studies, Indian Institute of Technology, Delhi, India, pp. 344.
[21] Sachs, N. W., 1991 "What Does PM Mean? Preventive vs. Predictive Maintenance," Lubr. Eng., pp. 889–891.
[22] Dong, W. P., Davis, E. J., Butler, D. L., and Stout, K. J., 1995, "Topographic Features of Cylinder Liners—An Application of Three-Dimensional Characterization Techniques," Tribol. Int., **28**(7), pp. 453–463.

A New Criterion to Determine the Start of Combustion in Diesel Engines

Tomaž Katrašnik
e-mail: tomaz.katrasnik@fs.uni-lj.si

Ferdinand Trenc

Samuel Rodman Oprešnik

Faculty of Mechanical Engineering,
University of Ljubljana,
Aškerčeva 6,
SI-1000 Ljubljana, Slovenia

A new criterion for the determination of the start of combustion (SOC) from the diesel engine in-cylinder pressure diagram was developed. It is defined as the maximum of the third-order derivative of the cylinder pressure with respect to the crank angle. This criterion declares SOC more precisely than other previously published criteria based on pressure diagnostics. This fact was proven analytically and was discernable from the analysis of the experimental data. Besides its accuracy it is also robust enough to allow automatic evaluation of the SOC during processing of the pressure data for a large number of cycles. By applying the first law of thermodynamics analysis to the engine cylinder it was discovered that the third-order derivative of the in-cylinder pressure with respect to the crank angle is the most suitable criterion for determination of the SOC from the in-cylinder pressure diagram. Subsequently, the criterion was validated through experimental data analysis of the in-cylinder pressure diagrams for various engine speeds and loads. In order to evaluate the rate of heat release (ROHR), which formed the base for the experimental validation, in-cylinder pressure diagrams were processed with a computer code based on the first law of thermodynamics. The cylinder pressure was measured with an advanced piezoelectric sensor at the resolution 0.1 deg CA. Top dead center was determined with the capacitive top dead center sensor. Due to the analytic foundation of the developed method and its validation through highly accurate experimental data it can be concluded that new criterion is credible for the determination of the SOC. [DOI: 10.1115/1.2179471]

Introduction

Selection of the proper ignition criterion is of crucial importance for a credible analysis of in-cylinder pressure diagrams, used for the determination of ignition delay period and evaluation of the ROHR. Definition of the SOC from the experimental data is still a complicated issue. The common term "ignition delay" or exactly "pressure rise delay" as proposed by Hardenberg and Hase [1] consists of physical delay and combined ignition delay of the fuel, which in turn consists of cold flame phase and blue flame phase [1,2]. As it is extremely difficult to distinguish between blue flame phase and the explosion flame phase [2], and taking into account that all these reactions are exothermic, there exists no exact-single point of the SOC. In addition, in the "ignition delay" period the exothermic preflame reactions are interlaced with the endothermic fuel droplet evaporation reactions. Therefore, the SOC is commonly defined by distinct pressure deviation from the pure compression pressure, which is the consequence of significant departure of the ROHR from 0.

The ignition criteria can roughly be divided into those based on pressure diagnostics [1,3–8] and those based on light emission [9,10]. Assanis et al. [7] stated that pressure-based techniques are more reliable than the light emission ones. Additionally, Heywood [8] indicates that the pressure change is often detected before the luminosity detector has noted the appearance of a flame, although Taylor [11] states that, on the average, data coincide well but are scattered. From the above arguments and the complexity of either of the above methods, pressure-based diagnostics can be obviously privileged.

The paper presents a new ignition criterion based on pressure diagnostics. Characteristics of published criteria will be discussed first. This chapter is followed by the analytic derivation of the new criterion, which was subsequently confirmed by highly accurate experimental data, the accordance of which is presented in the Results.

Literature Review

A suitable criterion for the accurate detection of the SOC should be independent of the engine load and speed. Such criterion is generally applicable to different engines operating at arbitrary running conditions. Criteria presented, in [3,4], which are defined by the rapid rise of the pressure, do not meet previously mentioned demands since they are speed and load dependent. Additionally, these criteria demand manual processing of the in-cylinder pressure data, which is unsuitable for the computer analysis of a large amount of the data. The criterion, which is defined as the point where pressure recovers to the value that corresponds to the compression without fuel evaporation and is presented in [5,6], is rather more suitable for bomb tests than for the analysis of internal combustion engine data. For the latter use a reference state must be computed, which introduces additional error and effort in processing the data.

Due to the similarity of the processes, criteria that have been used to characterize knock in spark ignition engines [12–15] are also suitable for detection of autoignition in diesel engines. Not all criteria [12–15] based on gradient and interval methods of pressure analysis are suitable for the determination of SOC since their primary task is knock detection and not its exact positioning. Criterion based on $\max(d^2Q_f/d\varphi^2)$ presented in [12,13] is similar to $\max(d^2p/d\varphi^2)$, which was used as the ignition criterion by [7,8]. Relation between these two criteria will be exposed later. This widely used criterion also predicts SOC with delay, also indicated in [12,13], and will be later proven analytically and confirmed with experimental data. Checkel and Dale [14] proposed the minimum of the third-order derivative of the cylinder pressure with respect to the crank angle as the appropriate criterion for knock

---

Contributed by the Internal Combustion Engine Division of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received January 14, 2005; final manuscript received September 26, 2005. Review conducted by D. Assanis.

detection. This criterion indeed clearly indicates the knock, but is unable to detect its exact starting point. Barton and co-authors [15] proposed the use of the pressure derivative to detect knock. This criterion shows the same weakness as already described in relation to [3,4] and is inaccurate as ignition criterion, which will be shown later.

Analytical Evaluation

In this chapter, analytic derivation of the original ignition criterion based on the maximum of the third-order derivative of the cylinder pressure with respect to the crank angle is presented.

After applying first law of thermodynamics to the control volume—cylinder—using ideal gas law

$$pV=mRT \tag{1}$$

and following relations for specific internal energy and gas constant

$$u=u(p,T,\lambda) \tag{2}$$

$$R=R(p,T,\lambda) \tag{3}$$

the ROHR is expressed as (see also [8,16–18])

$$\frac{dQ_f}{d\varphi}=H_d\frac{p\dfrac{dV}{d\varphi}\left(1+\dfrac{1}{RA}\dfrac{\partial u}{\partial T}\right)+m\left(\dfrac{\partial u}{\partial p}+\dfrac{\partial u}{\partial T}\dfrac{T}{p}\dfrac{B}{A}\right)\dfrac{dp}{d\varphi}-\dfrac{dQ_w}{d\varphi}}{H_d-u+\dfrac{T}{A}\dfrac{\partial u}{\partial T}+\dfrac{m_A}{\lambda_{st}m_f^2}m\left(\dfrac{\partial u}{\partial\lambda}-\dfrac{T}{R}\dfrac{\partial R}{\partial\lambda}\right)}, \tag{4}$$

where

$$A=1+\frac{T}{R}\frac{\partial R}{\partial T} \tag{5}$$

and

$$B=1-\frac{p}{R}\frac{\partial R}{\partial p}. \tag{6}$$

Taking into account temperature and pressure range in an internal combustion engine cylinder, relations (2) and (3) can be simplified to

$$u=u(T,\lambda) \tag{7}$$

and

$$R=R(\lambda) \tag{8}$$

without a noticeable loss of accuracy [18]. Equations (7) and (8) imply that $A=1$ and $B=1$, rewriting Eq. (4) to a simpler form

$$\frac{dQ_f}{d\varphi}=H_d\frac{p\dfrac{dV}{d\varphi}\left(1+\dfrac{1}{R}\dfrac{\partial u}{\partial T}\right)+\dfrac{V}{R}\dfrac{\partial u}{\partial T}\dfrac{dp}{d\varphi}-\dfrac{dQ_w}{d\varphi}}{H_d-u+T\dfrac{\partial u}{\partial T}+\dfrac{m_A}{\lambda_{st}m_f^2}m\left(\dfrac{\partial u}{\partial\lambda}-\dfrac{T}{R}\dfrac{\partial R}{\partial\lambda}\right)}. \tag{9}$$

For the analysis of SOC from the in-cylinder pressure diagrams it is reasonable to take into consideration only the significant terms and only the significant physical effects in Eq. (9). Considering the perfect gas simplification and omitting heat transfer to the walls, which is, compared to the other two terms of the numerator of Eq. (9), insignificant around the SOC, Eq. (9) can be rewritten as

$$\frac{dQ_f}{d\varphi}=\frac{1}{\kappa-1}\left(\kappa p\frac{dV}{d\varphi}+V\frac{dp}{d\varphi}\right). \tag{10}$$

In TDC Eq. (10) further simplifies to the form

$$\frac{dQ_f}{d\varphi}\propto\frac{dp}{d\varphi}. \tag{11}$$

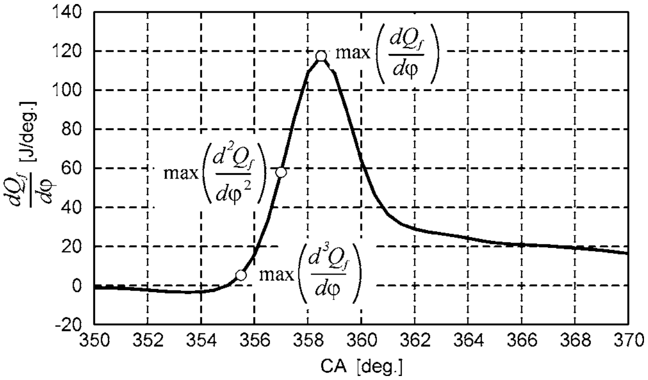


Fig. 1 ROHR with indicated maximum values of its derivatives

It will be shown later that relation (11) is valid also for the wider range around the TDC where SOC in diesel engines usually occurs, since the influence arising from the second term on the rhs of Eq. (10) is much stronger than that arising from the first one at the SOC (see also Fig. 4). It can therefore be concluded that in the vicinity of the SOC $dQ_f/d\varphi$ is related to $dp/d\varphi$. Thus, $\max(dp/d\varphi)$ coincides with $\max(dQ_f/d\varphi)$, which defines the maximum of the ROHR. This criterion obviously indicates SOC with delay. Following the same pattern $\max(d^2p/d\varphi^2)$ coincides with maximum increase of the ROHR ($\max(d^2Q_f/d\varphi^2)$). The latter criterion is evidently more suitable than the first one, but it still indicates SOC with delay; this can be deduced intuitively and is shown in Fig 1.

Intuitively and also physically adequate is to define SOC as the instant when $dQ_f/d\varphi$ significantly departs from 0, i.e., the point of maximum curvature of $dQ_f/d\varphi$ (Fig. 1: $\max(d^3Q_f/d\varphi^3)$), or, speaking mathematically, the maximum of the second-order derivative of $dQ_f/d\varphi$, which reads $\max(d^3Q_f/d\varphi^3)$. Following the idea presented above, the SOC from the diesel engine in-cylinder pressure diagram can be defined as

$$\varphi_{\mathrm{SOC}}=\max\left(\frac{d^3p}{d\varphi^3}\right). \tag{12}$$

The relevance of the above statements will be confirmed by the analysis of experimental in-cylinder pressure diagrams. The ROHR was derived from the in-cylinder pressure diagrams with the application of Eq. (4), using the author's FORTRAN code.

Experimental Setup

In the experiments a modern 6 cylinder, 6.87 liter turbocharged and aftercooled low emission truck diesel engine MAN D 0826 LOH 15, with 108/125 mm bore/stroke, compression ratio 18:1, equipped with the HOLSET H1E—8264BF/H16WA8 free-floating turbocharger and developing 162 kW/2400 rpm, was used. The laboratory experimental setup is capable of simultaneous measurements of engine system parameters as well as of cycle-resolved parameters. The engine is coupled with a Zöllner eddy-current dynamometer controlled by AVL BME-400 and Zöllner LSE 513. A 12 bit, 16 channel National Instruments data-acquisition system with maximum sampling frequency 1 MS/s was used for measurements of the in-cylinder pressure diagrams. A COM CAM UNIT Type 2613 shaft encoder provided an external trigger and an external clock (0.1–6 deg CA) for data-acquisition system. In-cylinder pressure was measured with calibrated piezo-electric pressure transducer AVL GH12D connected to calibrated charge amplifier COM PCA Type 2630. Maximum uncertainty of pressure measurement, which includes uncertainties of pressure transducer, charge amplifier, and A/D conversion, amounts to 0.31%. TDC was determined by capacitive sensor COM Type 2653 with maximum uncertainty less than 0.05 deg

CA (considering also engine speed change) whereas standard deviation of particular measurement was smaller than 0.006 deg CA. Maximum uncertainty of pressure measurements corresponding to crank angle was therefore 0.96%. The LabVIEW software was used to control the data-acquisition system. The LabVIEW environment was also applied for fast processing of input signals.

A large effort was devoted to the exact determination of TDC since its accuracy has a crucial influence on the ROHR. The output of the TDC sensor was therefore connected to the charge amplifier, which is generally applied for pressure measurements to diminish possible time delay of the signal when amplified by the charge amplifier. Only high pressure phase TDC signals from the capacitive sensor were processed since high and low pressure phase signals differ slightly and therefore increase standard deviation of the measurement.

The next important step towards better accuracy of the experimental data is accurate determination of pressure offset. This was done using two methods, based on two different principles. First, the inlet manifold pressure was used for precise pegging of the cylinder pressure, thus preventing the drifting of a signal. This approach, which also defines pressure offset, was additionally checked by a method based on the assumption of adiabatic behavior of the working medium during the compression stroke. A suitable interval where the heat transferred to working medium equals the heat rejected from the working medium must be chosen to minimize the error introduced by the above assumption. Many data from the in-cylinder pressure diagram, forming overdetermined system of equations, were then processed with computer code to evaluate the pressure offset. Agreement of the results of both methods assures us of the reliability of the in-cylinder pressure diagrams.

For subsequent analysis the average cycle obtained out of averaging 100 cycles was used as a representative cycle of the individual operating point of the engine. Very good coincidence of consecutive cycles proved the regularity of this approach, since with the averaging high frequency low amplitude errors introduced by the measuring equipment diminished significantly. Similarity of pressure amplitudes and their appropriate shift in the vicinity of the ignition point and throughout the entire combustion period of individual cycles proved the stability and repeatability of the measurements.

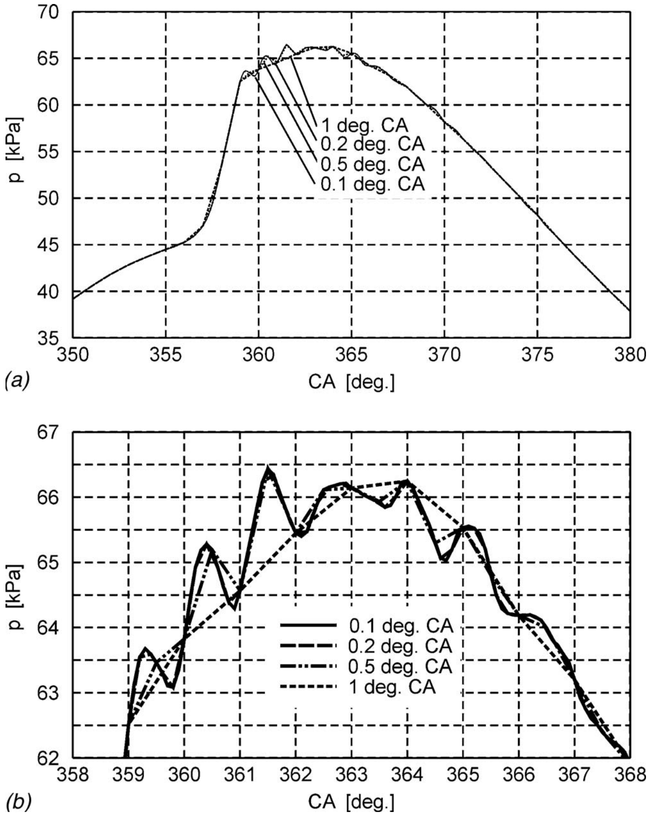


Fig. 2 In-cylinder pressure at 1200 rpm, 0.2 MPa bmep for different sampling resolutions

Results

Experimental data were further processed by the author's FORTRAN code, which performs the first law analysis of in-cylinder pressure data, data smoothing, and evaluation of the pressure derivatives. Numerical derivatives of first, second, and third order were evaluated by central schemes with fourth-order accuracy.

In Fig. 2 in-cylinder pressure diagrams for the operating conditions 1200 rpm, 0.2 MPa bmep, and sampling resolutions 0.1, 0.2, 0.5, and 1 deg CA are presented. Significant oscillations equivalent to predominant cylinder pressure resonance mode, determined by the configuration of the combustion chamber, are evident from Figs. 2(*a*) and 2(*b*) when high sampling frequencies are used. These pressure oscillations appear after the start of combustion and coincide perfectly with the calculated lowest excitation frequency of the gas in the combustion chamber, which corresponds to the first extreme of the Bessel function defined as $dJ_1(z)/dz=0$, for all operating conditions of the engine. Similar effects were observed by [12,14,15] when knock in SI engines was analyzed. It can be seen from Fig. 2(*b*) that pressure oscillations disappear at sampling resolution 1 deg CA, since sampling frequency is nearly equal to the primary gas vibration frequency. Accuracy of the pressure diagrams is assured due to the use of calibrated piezoelectric pressure transducer whose natural frequency is considerably above the sampling frequency for all sampling resolutions. Pressure oscillations observed at sampling resolution lower than 1 deg CA are therefore not the consequences of "pipe" effects or other measurement errors and, hence, adequately represent pressure at the measuring point. With the introduction of high sampling frequencies the measured pressure is no longer the representative in-cylinder pressure but only a local pressure. This difficulty could be overcome by use of multiple sensors placed at various radii and azimuth angles, since the averaged signal from all the sensors would be oscillation free. However, pressure oscillations appear a few deg CA after the start of combustion, as it was noted in Fig. 2(*a*), and therefore do not influence the determination of the SOC. At this place it should also be noted that pressure measured at 1 deg CA adequately represents the in-cylinder pressure, whereas pressure traces measured with higher sampling frequencies additionally carry imposed signals of cylinder pressure resonance modes, being also in accordance with the data presented in Refs. [12,14,15,19]. However, averaged data recorded at higher sampling frequencies coincide with the data recorded at 1 deg CA.

Pressure slope changes due to cylinder pressure resonance modes are larger than those due to the SOC when high sampling frequencies, capable of capturing cylinder pressure resonance oscillations, are used. Therefore, in the presented investigation the data recorded at sampling resolutions 0.1 and 0.2 deg CA were processed until the occurrence of these oscillations. On the other hand, the data recorded at sampling resolutions 0.5 and 1 deg CA were processed in the whole, since the derivatives of the cylinder pressure resonance oscillations have smaller magnitude than those of the SOC. The procedure used with sampling resolutions 0.1 and 0.2 deg CA is by no means incorrect, since the recorded pressure is no more the representative in-cylinder pressure but only a local pressure. This difficulty could be overcome, as already mentioned, by the use of many sensors, since differentiation of the averaged oscillation free data could be performed without restriction. The

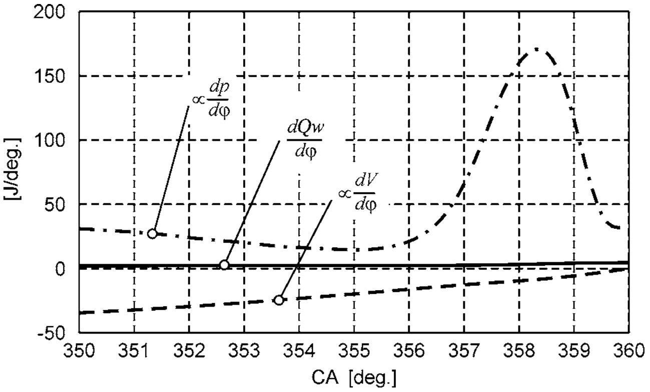


Fig. 3 The magnitude of the numerator terms in Eq. (4) at 1200 rpm, 0.2 MPa bmep

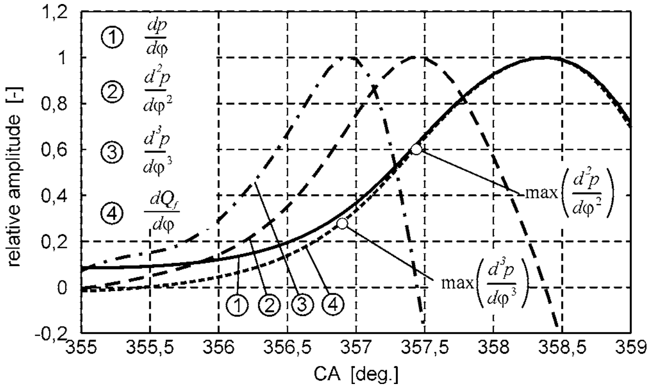


Fig. 5 Pressure derivatives and ROHR at 1200 rpm, 0.2 MPa bmep

high sampling frequency data were used to confirm the validity of the new SOC criterion, which is also applicable to data recorded with coarser sampling resolution.

At high sampling frequencies special care should be taken when preparing the data for postprocession. Frequency filtering during the measurement is not appropriate due to the shift of the data. Postprocession of the data with the Fourier transformation is appropriate for smoothing the pressure data, but the higher pressure derivatives start to oscillate with cutoff frequency. Data smoothing with weighted average of neighboring points turned out to be the most suitable smoothing technique for the necessity of the presented analysis. This technique was used with data measured at sampling resolutions of 0.1 deg CA, whereas almost all data recorded at 0.2 deg CA and all data recorded at 0.5 and 1 deg CA could be processed without smoothing. Furthermore, the results obtained when processing high accuracy data recorded at 0.1 deg CA resolution will be shown, and at the end of this chapter the results obtained when processing the data at different sampling resolutions will be compared.

Figure 3 represents the magnitude of the numerator terms in Eq. (4) for the engine running conditions: 1200 rpm, 0.2 MPa bmep (SOC 356.9 deg CA ATDC according to the proposed method). As already stated, it is obvious that the heat transfer term, calculated with the empirical formulas proposed by Hohenberg [20], is much smaller than the other two terms and could therefore be omitted in the derivation of Eq. (12).

Furthermore, Fig. 4 presents the magnitude of the terms obtained after twofold differentiation of the Eq. (10) for the operating conditions 1200 rpm, 0.2 MPa bmep. These results confirm the above statement that $\max(d^3Q_f/d\varphi^3)$ coincides with $\max(d^3p/d\varphi^3)$, which is indispensable in deriving Eq. (12). It can be observed that each term containing higher derivative of the pressure is nearly two orders of magnitude larger than the term containing the lower one in the vicinity of SOC.

As already stated in the Introduction it is difficult to define the exact position of the SOC since some exothermic reactions take place before the commonly defined SOC, i.e., distinct pressure deviation from the compression pressure, the latter being the consequence of significant departure of the ROHR from 0. Figure 5 presents $dp/d\varphi$, $d^2p/d\varphi^2$, $d^3p/d\varphi^3$, and $dQ_f/d\varphi$ for the operating conditions 1200 rpm, 0.2 MPa bmep. It is obvious that $\max(dp/d\varphi)$ coincides perfectly with $\max(dQ_f/d\varphi)$, as stated in the section "Analytical Evaluation." Furthermore, it can be seen that commonly used ignition criterion $\max(d^2p/d\varphi^2)$ coincides with maximum increase of the ROHR and therefore indicates SOC with delay, which again confirms the derivation presented in "Analytical Evaluation." From Fig. 5 it is also discernible that $\max(d^3p/d\varphi^3)$ precisely indicates the point of maximum curvature of $dQ_f/d\varphi$, which was the main request on the new SOC criterion.

Analyzing Fig. 5, it can be concluded that the point of maximum curvature of $dQ_f/d\varphi$ takes place slightly after the significant departure of the ROHR from 0. This discrepancy is the consequence of a complex set of physical and chemical reactions occurring during the "ignition delay" period, due to the occurrence of exothermic reactions prior to the explosion flame [2]. The delay between the significant departure of the ROHR from 0 and the point of maximum curvature of $dQ_f/d\varphi$ is not constant but depends on operating conditions (Figs. 5–7) and engine design. It is therefore still reasonable to define the SOC as the point of maximum curvature of $dQ_f/d\varphi$, since it has logical physical background and can somehow be related to the onset of explosion

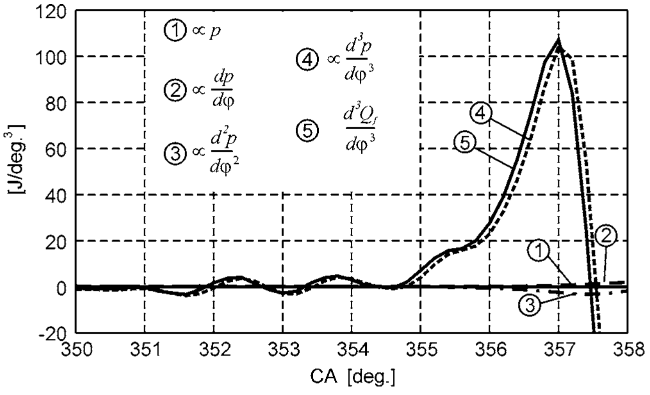


Fig. 4 The magnitude of terms obtained after twofold differentiation of Eq. (10) at 1200 rpm, 0.2 MPa bmep

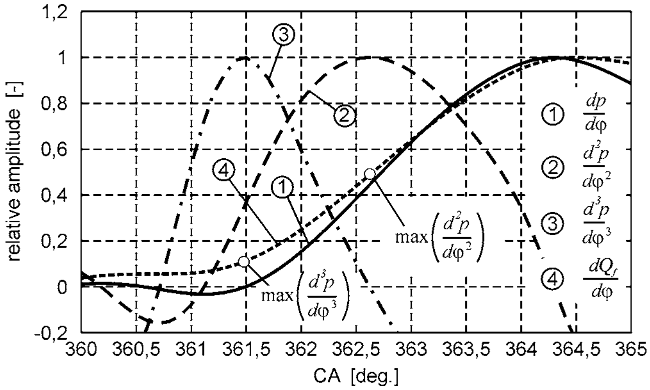


Fig. 6 Pressure derivatives and ROHR at 2100 rpm, 0.1 MPa bmep

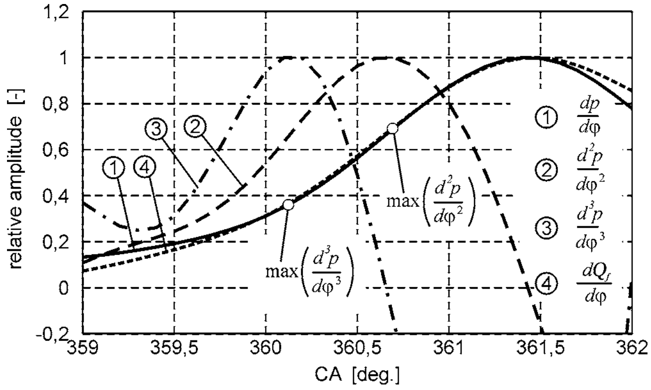


Fig. 7 Pressure derivatives and ROHR at 1500 rpm, 1.2 MPa bmep

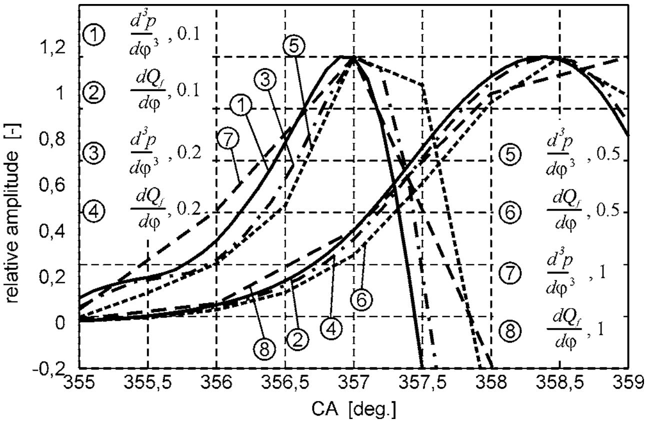


Fig. 8 Third-order derivative and ROHR for different sampling resolutions at 1200 rpm, 0.2 MPa bmep

flame. Another evidence of suitability of the presented criterion can also be found in Fig. 6, where the point of maximum curvature of $dQ_f/d\varphi$ coincides very well with the significant departure of the ROHR from 0.

The conclusions drawn for pressure derivatives in relation to Fig. 5 are also valid for Figs. 6 and 7. From Figs. 5–7 it is evident, considering Figs. 1 and 4, that $\max(d^3p/d\varphi^3)$ determines SOC more precisely and earlier than the $\max(d^2p/d\varphi^2)$: 0.5 deg CA at 1200 rpm, 0.2 MPa bmep; 0.6 deg CA at 1500 rpm, 1.2 MPa bmep; and 1.1 deg CA at 2100 rpm, 0.1 MPa bmep.

The observed operating conditions presented in Figs. 5–7 cover the broad engine operating range, necessary to validate the credibility of the new SOC criterion. Altogether, 11 engine operating points spread over the entire operating range were analyzed for the purpose of the presented study. It was found out that $\max(d^3p/d\varphi^3)$ defines SOC more precisely than the other two criteria based on pressure derivatives. $\max(d^3p/d\varphi^3)$ indicates SOC 0.4–1.3 deg CA earlier than the commonly used criterion $\max(d^2p/d\varphi^2)$ for the observed operating range of the investigated engine as shown in Table 1. In order to further validate accuracy and credibility of the new criterion, in-cylinder pressure diagrams of conceptually different STEYR Motors marine engine (4 cylinder, 2.1 liter high speed turbocharged diesel engine developing 70 kW at 4300 rpm) were also processed. The analysis of the results leads to the same conclusions, which were presented when analyzing the MAN engine data: $\max(d^3p/d\varphi^3)$ indicates SOC 0.6–1.6 deg CA earlier than $\max(d^2p/d\varphi^2)$ for the analyzed operating range, whereas SOC occurs in the interval between 356 and 366 deg CA. Hence, validity of the new criterion was additionally verified. The difference between $\max(d^3p/d\varphi^3)$ and $\max(d^2p/d\varphi^2)$ has great impact on the accuracy of the "ignition delay" period and ROHR determination, especially at low engine loads where combustion lasts only a few deg CA.

Table 1 $\max(d^3p/d\varphi^3)$, $\max(d^2p/d\varphi^2)$ and their difference for analyzed operating points

Engine Speed (rpm)	bmep (MPa)	max ($d^3p/d\varphi^3$) (deg CA)	max ($d^2p/d\varphi^2$) (deg CA)	Δ (deg CA)
1200	0.2	356.9	357.4	0.5
1200	0.8	356.7	358	1.3
1500	0.2	358.3	359	0.7
1500	0.8	358.5	359	0.5
1500	1.2	360.1	360.7	0.6
1800	0.1	361.2	361.7	0.5
1800	0.2	360.7	361.2	0.5
1800	0.8	360.3	360.8	0.5
2100	0.1	361.5	362.6	1.1
2100	0.2	361.2	361.9	0.7
2100	0.8	361.6	362	0.4

All results presented in Figs. 3–7 were derived from in-cylinder pressure data measured at sampling resolutions 0.1 deg CA. Comparison of $d^3p/d\varphi^3$ and $dQ_f/d\varphi$ for sampling resolutions 0.1, 0.2, 0.5, and 1 deg CA is presented in Fig. 8. The results (positions of the peaks) coincide very well. The differences in $\max(d^3p/d\varphi^3)$ and $\max(dQ_f/d\varphi)$ for different sampling resolutions are below the particular sampling resolution. This fact corroborates the confidence in using the new SOC criterion also on the data recorded with coarser sampling resolution. Even when the difference between $\max(d^3p/d\varphi^3)$ and $\max(d^2p/d\varphi^2)$ is between 0.5 and 1 deg CA at the sampling resolution 0.1 deg CA, the application of these two criterions on the data recorded at the sampling resolution 1 deg CA predicts the difference of 1 deg CA.

Figures 9 and 10 presents $dp/d\varphi$, $d^2p/d\varphi^2$, $d^3p/d\varphi^3$, and $dQ_f/d\varphi$ for the operating conditions equivalent to that in Fig. 5 for sampling resolutions 0.5 and 1 deg CA. All conclusions valid for the sampling resolution 0.1 deg CA are also valid for the coarser sampling resolutions. Comparing Figs. 9 and 10 it can be seen that the magnitude of the derivatives of the cylinder pressure resonance oscillations (Fig. 2(*b*)) diminish with the coarser sampling resolutions. An additional peak of $d^3p/d\varphi^3$ occurring after $\max(dQ_f/d\varphi)$ can be noticed for some operating conditions. This difficulty in the definition of the SOC with $\max(d^3p/d\varphi^3)$ could be overcome when a $\max(d^3p/d\varphi^3)$ prior to $\max(d^2p/d\varphi^2)$ is considered. This physically intuitive rule also confirmed its robustness when analyzing various in-cylinder pressure data.

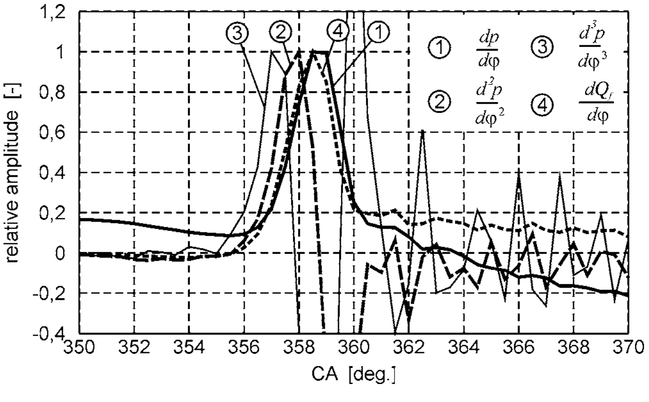


Fig. 9 Pressure derivatives and ROHR for sampling resolution 0.5 deg CA at 1200 rpm, 0.2 MPa bmep

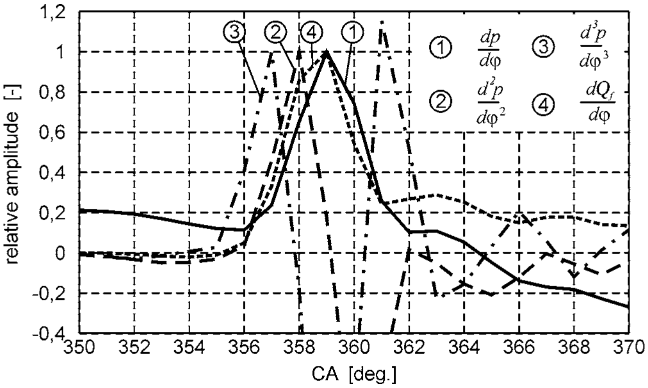


Fig. 10 Pressure derivatives and ROHR for sampling resolution 1 deg CA at 1200 rpm, 0.2 MPa bmep

Conclusion

A new accurate criterion for determination of the SOC from the diesel engine in-cylinder pressure diagram was developed and validated. The criterion has a clear physical background, which was confirmed with many results obtained from high accuracy experimental data recorded for various engine operating regimes. The new criterion defines SOC with higher precision than other previously published criteria based on pressure diagnostics. The criterion is robust enough to allow automatic evaluation of the SOC during processing of the pressure data for a large number of cycles, since its robustness is in the range of the criterion based on the second derivative. When processing high sampling frequency data care should be taken, since cylinder pressure resonance oscillations could disfigure the results. However, this is not due to the deficiency of the criterion but to faultiness in the representation of the measured data, since the physical relevance of the experimental data should not be lost out of focus, as it was explained above.

Acknowledgment

The financial support of the Ministry of Education, Science and Sports of the Republic of Slovenia is greatly appreciated. Valuable support of the MAN Slovenia is also gratefully acknowledged.

Nomenclature

Symbols

H_d = lower heating value of the fuel, kJ/kg
J = Bessel function of the first kind
m = mass, kg
p = pressure, Pa
Q = heat, J
R = gas constant, kJ/kgK
T = temperature, K
u = specific internal energy, m^2/s^2
V = volume, m^3
φ = crank angle, deg CA
κ = isentropic exponent
λ = air-fuel ratio

Subscripts

A = air
f = fuel
st = stoichiometric
w = wall

Abbreviations

A = after; referring to TDC of low pressure phase
bmep = brake mean effective pressure
CA = crank angle
ROHR = rate of heat release
SOC = start of combustion
TDC = top dead center

References

[1] Hardenberg, H. O., and Hase, F. W., 1979, "An Empirical Formula for Computing the Pressure Rise Delay of a Fuel From its Cetane Number and From the Relevant Parameters of Direct-Injection Diesel Engines," SAE Paper No. 790493.
[2] Sitkei, G., 1963, "Über den dieselmotorischen Zündverzug," Motortech. Z., **24**(6), pp. 190–194.
[3] Wong, C. L., and Steere, D. E., 1982, "The Effect of Diesel Fuel Properties and Engine Operating Conditions on Ignition Delay," SAE Paper No. 821231.
[4] Xia, Y. Q., and Flanagan, R. C., 1987, "Ignition Delay—A General Engine/Fuel Model," SAE Paper No. 870591.
[5] Aligrot, C., Champoussin, J. C., Guerrassi, N., and Claus, G., 1997, "A Correlative Model to Predict Autoignition Delay of Diesel Fuels," SAE Paper No. 970638.
[6] Kobori, S., Kamimoto, T., and Aradi, A. A., 2000, "A Study of Ignition Delay of Diesel Fuel Sprays," Int. J. Engine Res., **1**(1), pp. 29–39.
[7] Assanis, D. N., Filipi, Z. S., Fiveland, S. B., and Syrimis, M., 2003, "A Predictive Ignition Delay Correlation Under Steady-State and Transient Operation of a Direct Injection Diesel Engine," ASME J. Eng. Gas Turbines Power, **125**, pp. 450–457.
[8] Heywood, J. B., 1988, *Internal Combustion Engine Fundamentals*, Mc-Graw-Hill, New York.
[9] Pischinger, F., Reuter, U., and Scheid, E., 1988, "Self-Ignition of Diesel Sprays and Its Dependence on Fuel Properties and Injection Parameters," ASME J. Eng. Gas Turbines Power, **110**, pp. 399–404.
[10] Cowell, L. H., and Lefebvre, A. H., 1986, "Influence of Pressure on Autoignition Characteristics of Gaseous Hydrocarbon-Air Mixtures," SAE Paper No. 860068.
[11] Taylor, C. F., 1985, *The Internal-Combustion Engine in Theory and Practice*, Volume 2, The M.I.T. Press, Cambridge, MA.
[12] Ando, H., Takemura, J., and Koujina, E., 1989, "A Knock Anticipating Strategy Basing on the Real-Time Combustion Mode Analysis," SAE Paper No. 890882.
[13] Michael, S., Shigahara, K., and Assanis, D. N., 1996, "Correlation Between Knock Intensity and Heat Transfer Under Light and Heavy Knocking Conditions in a Spark Ignition Engine," SAE Paper No. 960495.
[14] Checkel, M. D., and Dale, J. D., 1986, "Computerized Knock Detection from Engine Pressure Record," SAE Paper No. 860028.
[15] Barton, R. K., Lestz, S. S., and Duke, L. C., 1970, "Knock Intensity as a Function of Engine Rate of Pressure Change," SAE Paper No. 700061.
[16] Katrašnik, T., 2001, "Analysis of Transient Process in a Turbocharged Diesel Engine," Master of Science thesis No. M/1177, Dept. of Mech. Engineering, University of Ljubljana, Slovenia.
[17] Katrašnik, T., 2004, "Numerical Modeling of Transient Processes in a Turbocharged Diesel Engine," Doctor of Science thesis No. Dr/278, Dept. of Mech. Engineering, University of Ljubljana, Slovenia.
[18] Pischinger, R., Krassnig, G., Taučar, G., and Sams, T., 1989, *Thermodynamik der Verbrennungskraftmachine*, Springer-Verlag, Berlin.
[19] Walter, T., Brechbuechl, S., Wolfer, P., and Payerl, N., 2005, "Pressure Indicating With Measuring Spark Plugs—Status of Performance," 2005 EAEC European Automotive Congress, Paper No. EAEC05YU-EN11.
[20] Hohenberg, G., 1979, "Advanced Approaches for Heat Transfer Calculation," SAE Paper No. 790825.

Angelo Algieri
e-mail: a.algieri@unical.it

Sergio Bova
e-mail: s.bova@unical.it

Carmine De Bartolo
e-mail: c.debartolo@unical.it

Mechanics Department,
University of Calabria,
87030 Arcavacata di Rende (CS), Italy

Influence of Valve Lift and Throttle Angle on Intake Flow in a High-Performance Four-Stroke Motorcycle Engine

A high-performance four-stroke motorcycle engine was analyzed at a steady flow rig. The aim of the work was to characterize the fluid dynamic behavior of the engine head during the intake phase. To this purpose a twofold approach was adopted: the dimensionless flow coefficient was used to evaluate the global breathability of the intake system, while the laser doppler anemometry (LDA) technique was employed to define the flow structure within the combustion chamber. The analysis gave evidence of two contrarotating vortices with axes parallel to the cylinder axis and showed variations in the flow structure when moving away from the engine head. Furthermore, the study highlighted the great influence of the throttle angle on the head fluid dynamic efficiency and how this influence changes with the valve lift. Experimental data were correlated by a single curve adopting a new dimensionless plot. Moreover, LDA measurements were used to evaluate the angular momentum of the flux and an equivalent swirl coefficient, and to correlate them to a previous global swirl characterization carried out on the same engine head using an impulse swirl meter. [DOI: 10.1115/1.2180277]

Introduction

A profound understanding of the intake process is a fundamental key to developing and optimizing modern internal combustion engines (ICEs) [1–3]. Breathing efficiency, in fact, strongly influences both the engine performances and the efficiency of the combustion process. Engine power and torque, at a given speed, increase as the mass flow rate of the air entering the combustion chamber increases. As a consequence, research leads to extreme designs of the head configurations in high-performance engines to maximize volumetric efficiency. The distance between the intake valve and the cylinder wall and between intake valves (in multivalve engines) are often reduced, and complex valve-to-wall and valve-to-valve interference phenomena take place during the intake phase [4,5]. Furthermore, the in-cylinder flow field is a basic element because it determines the progress and the quality of the combustion process and, therefore, fuel economy and emissions. In particular, organized large-scale flow structures, such as tumble and swirl, are often required during the induction phase to improve combustion efficiency and fuel economy and to respect the ever more severe regulations concerning exhaust emissions [6–8].

Nowadays, a detailed characterization of the fluid dynamic behavior of internal combustion engines during intake becomes necessary to design, develop, and optimize new high-efficiency engines. To this purpose different investigation tools, based both on CFD codes [9–11] and experimental approaches [1,12,13] are available. In the last few years, a host of experimental investigations have been carried out adopting dimensionless discharge and flow coefficients with the aim to provide global information on engine head breathability [14–16], while laser doppler anemometry (LDA) [17,18], particle image velocimetry (PIV) [3,19], and hot wire anemometry (HWA) [2,20] techniques have been used to define the flow field inside the cylinder of internal combustion engines. In particular, steady flow testing is a widely adopted procedure in the ICE research community to study the fundamentals of the intake process and to evaluate the flow capacity and the in-cylinder flow field of the actual engines, owing to its relative simplicity and the proper simulation of the real intake phase [21–23].

The present paper aims to analyze the fluid dynamic behavior of a production high-performance motorcycle engine during the intake phase. The four-stroke, dual-spark engine has been examined at a steady flow rig in order to have detailed information both on the global volumetric efficiency of the engine and on the flow structure within the combustion chamber. The effect of the throttle valve opening on the head breathability was investigated by means of the flow coefficients. In fact, few quantitative studies on the influence of the throttling process on engine volumetric efficiency are available in literature. Moreover, the LDA technique was employed to characterize the flow pattern in the cylinder and to evaluate how the flow field varies moving away from the engine head. Finally, the LDA data were correlated to some previous swirl measurements carried out using an impulse swirl meter [24]. To this purpose the dimensionless swirl coefficient was considered.

Experimental Apparatus

The experimental investigation focused on the head of a high-performance four-stroke internal combustion engine. The head is part of an L-twin-cylinder motorcycle engine, characterized by two valves and two spark plugs per cylinder. Figure 1 shows the engine head, and Table 1 lists its main characteristics.

The experimental analysis was carried out by means of a steady-flow rig, enabling air to be forced through the intake system of the engine head by means of a blower, while the valve lift is fixed to a selected value. The flow rig can deliver flow rates in the range 40–600 m^3/hr for ambient-cylinder pressure drops up to about 10 kPa at low flow rates and 5 kPa at the highest flow rates.

In addition to global mass flow rate, the facility also enables local velocity measurements with LDA. To this purpose, optical access to the cylinder was obtained through a window perpendicular to the cylinder axis. The LDA system is a one-color system (i.e., capable of measuring one component of the velocity) in a

---

Contributed by the Internal Combustion Engine Division of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received April 27, 2005; final manuscript received January 9, 2006. Review conducted by J. C. Cowart.

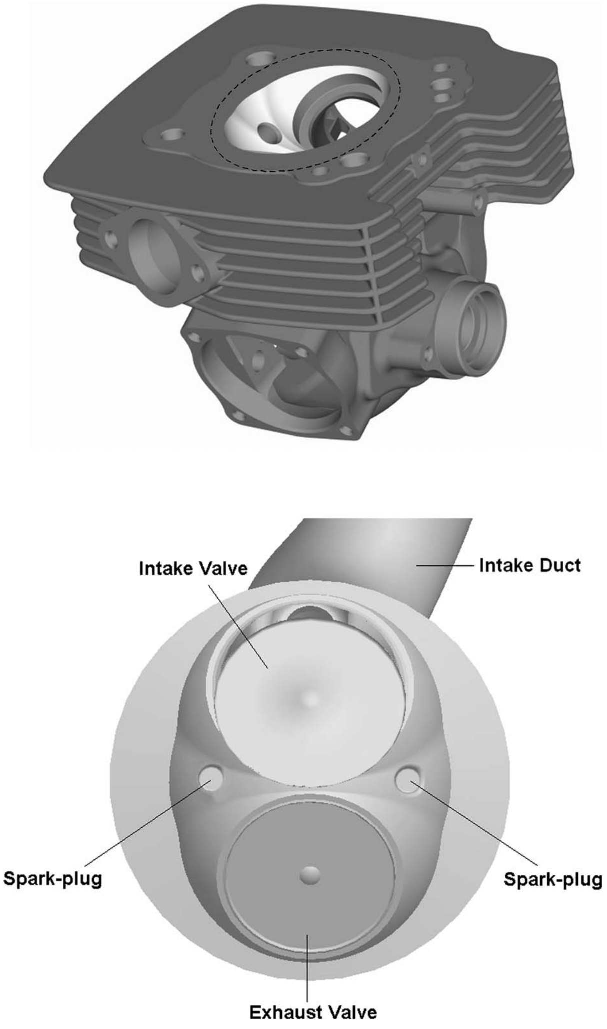


Fig. 1 Engine head

backscattering configuration, with a Bragg-cell frequency shifter. The system uses an argon-ion laser as light source (2W on the green line, at $\lambda = 514.5$ nm) and optical fibers for both transmitting and collecting optics. The main geometrical data of the optical system are:

- *beam spacing* 38 mm
- *focal length* 400 mm
- *probe volume width* 0.194 mm
- *probe volume length* 4.09 mm
- *number of fringes* 35
- *fringe spacing* 5.42 μm

If a frequency-shifter module is used, the number of fringes depends on other parameters, such as record length, center frequency, and bandwidth used by the signal processor. In the case of the reported measurements, the resulting number of fringes was 48.

Table 1 Characteristics of the engine

Stroke/Bore	L/B	0.761
Intake valve diameter/Bore	D_v/B	0.479
Throttle diameter/Bore	D_t/B	0.479
Compression ratio	r_c	10
Engine power		69.2 kW at 8000 rpm
Engine torque		94.2 Nm at 6000 rpm
Cooling system		Air cooled

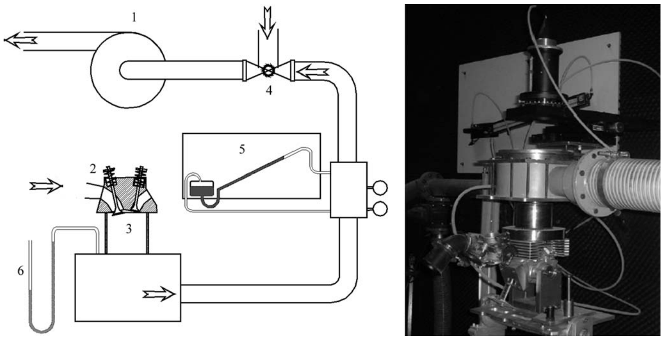


Fig. 2 Experimental setup

The movement of the LDA probe is obtained by using a micrometer *x-y* traversing system. The probe can also be rotated around its axis and moved vertically (*z*-axis). A dedicated signal processor, which performs fast-Fourier transform processing of the original signal in order to extract the Doppler frequency, is used for the analysis of the Doppler signal and ensures rejection of the signal produced by different particles that may be present within the measuring volume at the same time. The processor is linked to a computer in order to store and analyze the data.

For the seeding system, a fluidized bedlike scheme was employed. A fraction of the inlet air is first dried and then passed through a horizontal porous diaphragm on the top of which a layer of silica "microballoons" is deposited. The airstream then carries the particles and is subsequently mixed with the main intake air at the engine inlet. Figure 2 shows a scheme and a picture of the experimental set-up; more details have been given in [25–27].

Experimental Investigation

A twofold approach was adopted to analyze the fluid dynamic behavior of the high-performance four-stroke motorcycle engine during the intake phase. The engine head was examined in terms of global performances and in terms of local velocity measurements. In particular, the flow coefficient was used to define the head breathability, while the laser doppler anemometry (LDA) technique was employed to determine the flow field inside the combustion chamber.

Flow Coefficient. The usual flow coefficient was used as global performance indicator [1,12]. It is defined as a ratio of the actual (measured) mass flow rate to reference mass flow rate

$$C_f = \frac{\dot{m}_{\mathrm{meas}}}{\dot{m}_r} \tag{1}$$

If p_0 and T_0 are the ambient pressure and temperature and p_C is the cylinder pressure, and the flow is subsonic, them the reference mass flow rate is given by

$$\dot{m}_r = A_{\mathrm{ref}} \frac{p_0}{\sqrt{RT_0}} \left(\frac{p_C}{p_0}\right)^{1/\gamma} \left\{ \frac{2\gamma}{\gamma - 1} \left[1 - \left(\frac{p_C}{p_0}\right)^{(\gamma-1)/\gamma} \right] \right\}^{1/2} \tag{2}$$

if, instead, the flow is choked, then the mass flow is formalized as follows:

$$\dot{m}_r = A_{\mathrm{ref}} \frac{p_0}{\sqrt{RT_0}} \gamma^{1/2} \left(\frac{2}{\gamma+1}\right)^{(\gamma+1)/[2(\gamma-1)]} \tag{3}$$

The reference area is the valve outer seat area

$$A_{\mathrm{ref}} = \frac{\pi D_v^2}{4} \tag{4}$$

Table 2 Global analysis: measuring conditions

		First experimental campaign	Second experimental campaign
Pressure drop (kPa)	Δp	7.3	7.3
Dimensionless valve lift $(-)$	L_v/D_v	0.056–0.259 (step 0.0185)	0.093, 0.148, 0.178, 0.259
Throttle angle (deg)	F	45–85	45–85
Throttle step (deg)	ΔF	10	5

where D_v is the intake valve diameter.

Moreover, an absolute flow coefficient C_{abs} was defined to characterize the intake system efficiency independently from the valve lift [29]

$$C_{\text{abs}} = \frac{\phi_{Am}}{\phi_{Ad}} \tag{5}$$

ϕ_{Ad} represents the dimensionless theoretical flow rate downstream of the intake valve, based on isentropic flow condition

$$\phi_{Ad} = \sqrt{\frac{2}{\gamma - 1}\left[\left(\frac{p_C}{p_0}\right)^{2/\gamma} - \left(\frac{p_C}{p_0}\right)^{(\gamma+1/\gamma)}\right]} \tag{6}$$

while ϕ_{Am} is the dimensionless actual flow rate, averaged over the dimensionless valve lift

$$\phi_{Am} = \frac{\displaystyle\int_0^{(L_v/D_v)_{\max}} \frac{\dot{m}_{\text{meas}}}{A_{\text{ref}}\rho_0 a_0} d\frac{L_v}{D_v}}{\left(\dfrac{L_v}{D_v}\right)_{\max}} \tag{7}$$

where L_v is the valve lift, ρ_0 is the air density at the inlet, and a_0 is the sound speed.

The global measurements were carried out in two experimental campaigns for different valve lift and throttle angles. The ambient-cylinder pressure drop was set at $\Delta p = 7.3$ kPa. This value makes sure that the intake flow is fully turbulent and that the dimensionless flow coefficients are independent of the pressure drop for all the analyzed configurations. In fact, to this purpose the "port Reynolds number"[1] should always exceed 60,000 at low valve lifts and 90000 at high valve lifts [12] and this requirement is fully respected in this work. Table 2 summarizes the analyzed configurations.

The overall uncertainty of flow coefficients and absolute flow coefficients was always lower than 3%, and it decreased with valve lift and throttle angle. In particular, for $L_v/D_v > 0.074$ the uncertainty dropped to a value that was lower than 2%.

LDA Measurements. The local characterization of the flow inside the combustion chamber was achieved by the laser doppler anemometry technique. The measurements were carried out fixing the valve lift ($L_v/D_v = 0.178$) and the ambient-cylinder pressure drop ($\Delta p = 7.3$ kPa). The wide-open-throttle (WOT) configuration was considered. Data were collected on three measuring planes perpendicular to the cylinder axis at different distances from the engine head. Figure 3(*a*) shows the measuring planes and the relative distance from the head ($H/B = 1/4$, $H/B = 1/2$, and $H/B = 3/4$). B is the cylinder bore.

Thirty-three measuring points, distributed on four equally spaced diameters (D_1, D_2, D_3, and D_4), were defined on each measuring plane (Fig. 3(*b*)). In particular, the D_3 diameter is rotated by 5 deg clockwise with respect to the engine head symmetry axis (y-axis). At each measuring point, two velocity components (along the x and y directions) were recorded. To verify the previous measurements, a third component was acquired and compared to the value calculated from the x and y components, recording a good agreement. Three thousand samples were collected for each velocity component. The typical data rate was 0.1–0.3 kHz with 70–95% of the data valid. The relative uncertainty on the LDA measurements, which is mainly due to the setup of the electronic system, was lower than 2.2%.

Furthermore, the LDA measurements were used to evaluate the overall angular momentum of the flux M with respect to the cylinder axis on each measuring plane. To this purpose, the angular momentum produced by the tangential velocity v_t was combined with the axial mass flow, as [2,7,30]

$$M = 2\pi \int_0^{B/2} \rho v_a v_t r^2 dr \tag{8}$$

where v_a is the axial velocity, v_t is the tangential velocity, r is the distance of the generic measuring point from the cylinder axis, and ρ is the air density.

Finally, an equivalent swirl coefficient C_s was defined as the ratio of the flow angular to axial momentum [2,12]

$$C_s = \frac{8M}{\dot{m}_{\text{meas}} v_0 B} \tag{9}$$

where v_0 is a reference velocity, calculated from the isentropic relation for flow from a converging nozzle emptying into a plenum

$$v_0 = \left\{\frac{2\gamma}{\gamma - 1}\frac{p_0}{\rho_0}\left[1 - \left(\frac{p_c}{p_0}\right)\right]^{(\gamma-1)/\gamma}\right\}^{1/2} \tag{10}$$

The overall uncertainty of the angular momentum of the flux and swirl coefficient was always lower than 3.0% and 3.5%, respectively.

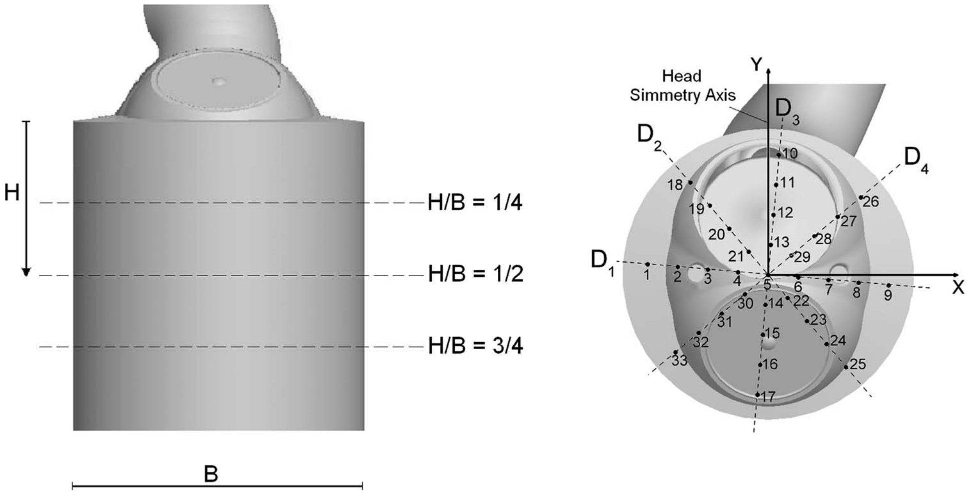
Results

Figure 4 shows the fluid dynamic efficiency of the engine head in terms of flow coefficient as a function of the dimensionless valve lift (L_v/D_v). The figure refers to the WOT configuration. A progressive increase in the flow coefficient is observed when the valve lift increases. This is due to an upsurge in the mass flow rate entering the combustion chamber until $L_v/D_v \leqslant 0.222$. A plateau is reached for $L_v/D_v > 0.222$, and there are negligible effects on the head breathability owing to the dimensions of the intake port and of the valve stem that define the minimum flow area at high valve lifts. Figure 4 depicts also the presence of three different regions characterized by decreasing slopes in the flow coefficient's curve that corresponds to different flow regimes, in line with the results found in the literature [1,31]. In particular, for low valve lift, the flow remains attached to the valve head and seat due to the high viscous phenomena. While raising the curtain area, a flow separation occurs, first of all, at the valve head and, successively, at the valve seat.

In order to investigate the influence of throttle valve opening on head breathability, the global measurements were repeated at various throttle angles. The analyzed configurations are reported in Table 2.

---

[1] The port Reynolds number is defined as the Reynolds number calculated at the intake port. More details are given in [12,28].



a) b)

Fig. 3 LDA analysis: measurement locations

Figure 5 shows similar behaviors for the different conditions and the presence of the three flow regimes. It is possible to observe that the "transition" from a flow condition to another one is reached at lower valve-lift values when the flow is throttled. As an example, the first transition phenomena for the WOT configuration ($F=85$ deg) occur at the valve lift $L_v/D_v=0.115$, whereas for $F=45$ deg they develop at $L_v/D_v=0.085$. In addition, the plot illustrates the noticeable influence of the throttle position on engine volumetric efficiency. Head breathability increases progressively with throttle opening for $F\leqslant 75$ deg and becomes constant for $F>75$ deg. This trend is plain to see in Table 3, which shows the absolute coefficients (calculated according to Eq. (5)) for the five throttle angles. Furthermore, Table 3 highlights that the relative increase in head breathability diminishes when the throttle valve is open more. In fact, the absolute flow coefficient raises by 16.1% when F passes from 45 deg to 55 deg, whereas the percentage drops to 6.4% moving from $F=65$ deg to $F=75$ deg, and finally, it becomes negligible for $F>75$ deg.

The effect of the throttle angle on the flow coefficients was investigated more closely at four valve lifts ($L_v/D_v=0.093$, 0.148, 0.178, and 0.259). To this purpose the throttle step was reduced from $\Delta F=10$ deg to $\Delta F=5$ deg.

Figure 6 confirms the previous trends and illustrates that the experimental data are well approximated by parabolic curves (dashed lines) and that also the absolute flow coefficients, shown in Table 3, lie on a parabola. Furthermore, the diagram displays, for each valve lift, the flow coefficient variations ΔC_f between the two extreme throttle positions ($F=45$ deg and $F=85$ deg). It is clear that the higher the valve curtain, the larger the influence of the throttle valve. All the experimental data are resumed in the contour plot of Fig. 7.

A systematic data analysis allowed the definition of a new in-

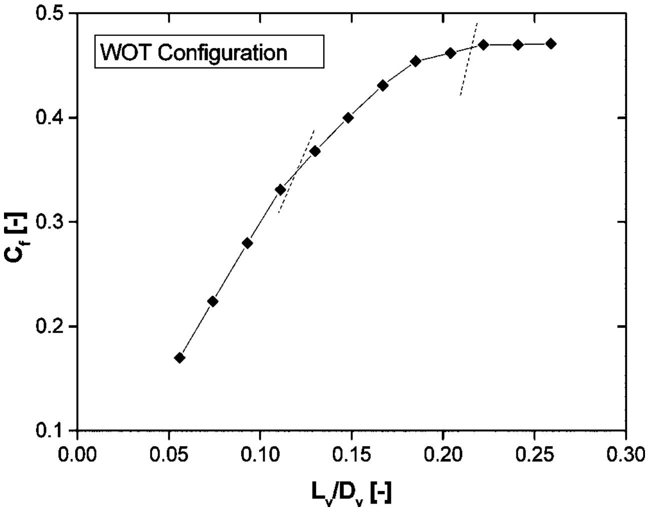


Fig. 4 Effect of the valve lift on head breathability, WOT configuration

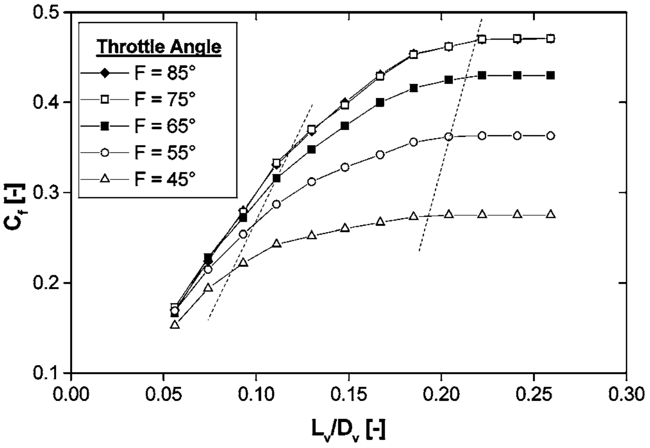


Fig. 5 Effect of the throttle angle on head breathability

Table 3 Absolute coefficient and relative percentage variation

F (deg)	C_{abs} (−)	ΔC_{abs} (%)
45	0.238	
55	0.296	16.1
65	0.337	11.4
75	0.360	6.4
85	0.360	0.0

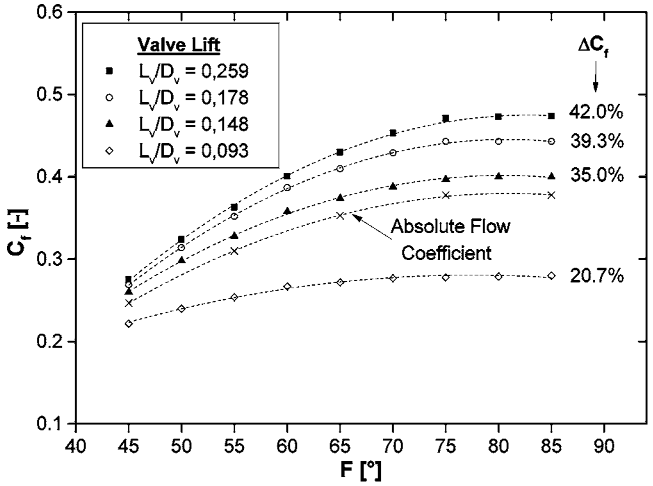


Fig. 6 Influence of the throttle angle on head breathability for four dimensionless valve lift

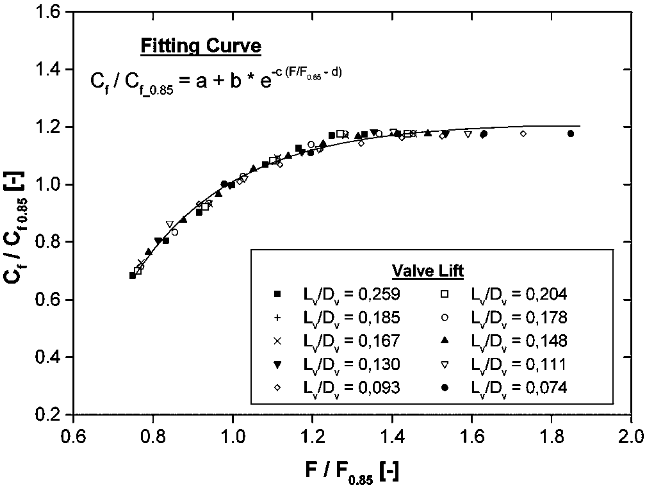


Fig. 8 Unique trend in the flow coefficient

teresting way to represent the flow coefficients. Experimental data were correlated by a single curve if they were plotted as

$$\frac{C_f}{C_{f_{0.85}}} = f\left(\frac{F}{F_{0.85}}\right) \tag{11}$$

where $C_{f_{0.85}}$ is the flow coefficient corresponding to 85% of maximum flow coefficient and $F_{0.85}$ is the throttle angle where the flow coefficient is equal to the 85% of maximum flow coefficient.

Figure 8 portrays, in fact, that all the experimental data tend to distribute along the same curve apart from the analyzed configuration of the head. In particular, the data are well approximated by an exponential curve, whose functional form is

$$\frac{C_f}{C_{f_{0.85}}} = (a+b)e^{-c[(F/F_{0.85})-d]} \tag{12}$$

where the values of the four constants are $a=1.2135$, $b=-0.5400$, $c=3.8835$, and $d=0.7489$. The result is very useful because it permits a drastic reduction in the measurements that have to be made to characterize the engine breathability at different valve lifts and different loads, reducing, as a consequence, time and costs of the investigations. Moreover, plotting the experimental data in the suggested form facilitates the check during the measuring phase.

The applicability of the previous rule to other engines was verified by considering a production four-cylinder automobile spark-ignition engine. The analysis showed the existence of a universal trend in the head breathability also for this engine. Specifically, it is interesting to observe (Fig. 9) that the distributions referring to the two engines tend to be very similar, despite the high differences in the engine geometry and characteristics and in the head breathability. More details are given by the authors in [32].

To examine the fluid dynamic behavior of the engine head in more detail and to characterize the flow field inside the combustion chamber, the LDA technique was used. Measurements were carried out by fixing the valve lift ratio at $L_v/D_v=0.178$ in the WOT configuration and by considering three measuring planes perpendicular to the cylinder axis, as illustrated in Fig. 3. Figure 10 depicts the mean velocities measured on the four diameters. Data refer to the intermediate plane, half bore from the engine head ($H/B=1/2$).

The velocity vectors give evidence of the existence of two contrarotating vortices whose axes are parallel to the cylinder axis. In the central region, the flow tends to distribute along the y-axis. The directions of the velocity vectors at the measuring locations 4–6 and 11–16 correspond, in fact, to the y direction. The center of the combustion chamber is also characterized by the largest

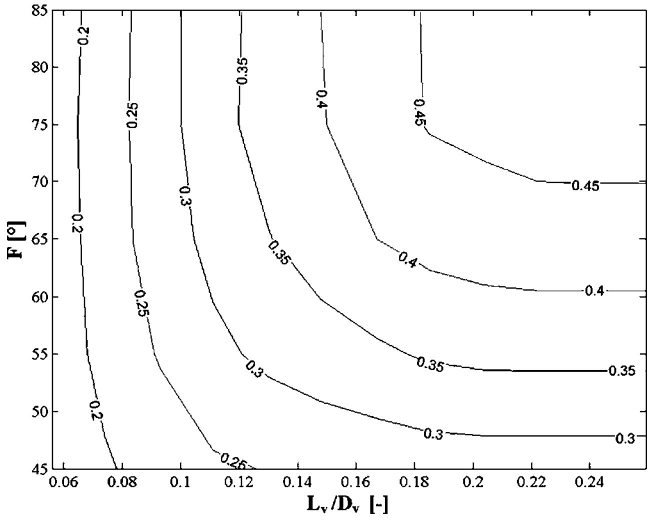


Fig. 7 Flow coefficient contour plot

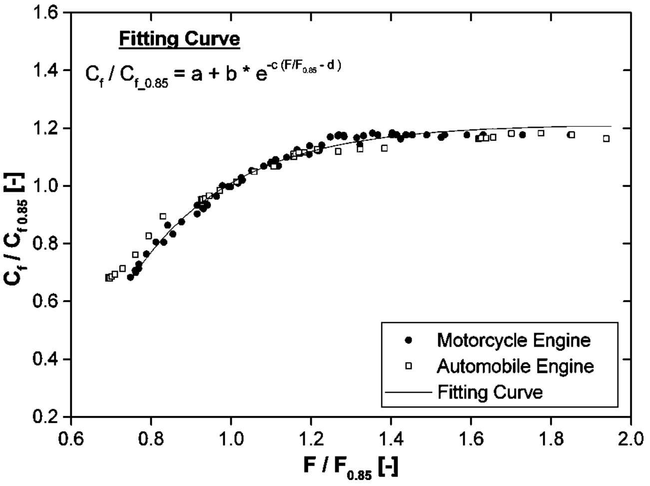


Fig. 9 Dimensionless plot for different engines

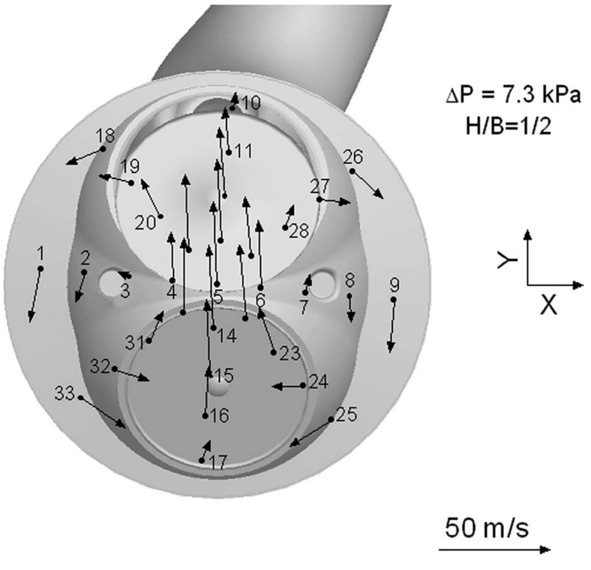


Fig. 10 Velocity vectors on the middle plane ($H/B = 1/2$)

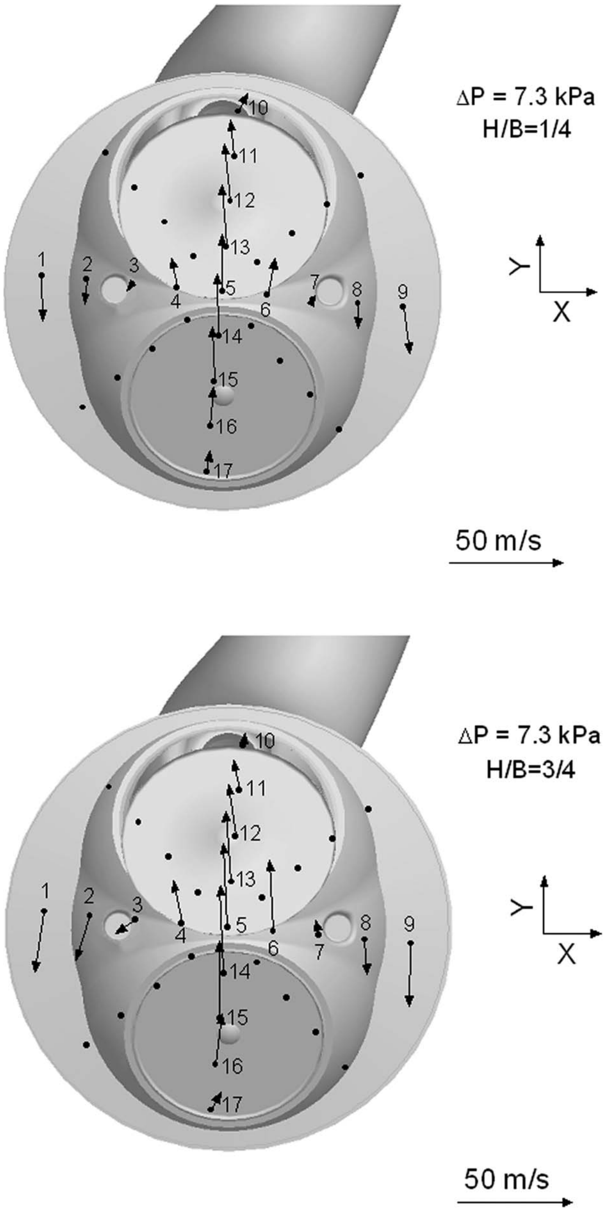


Fig. 11 Velocity vectors on lower and higher measuring plane

velocity values, while the lowest velocities are localized near to the cylinder wall (points 10 and 17) and near to the rotational axes of the two vortices (points 2–3 and 7–8). Furthermore, LDA measurements show a slight asymmetry in the flow-field distribution due to the position of the intake duct.

The local analysis was repeated for the other two measuring planes ($H/B = 1/4$ and $H/B = 3/4$) considering the same engine configuration ($L_v/D_v = 0.178$ and WOT). The measurements were restricted to D_1 and D_3 diameters. LDA data depict similar flow fields on the three planes with the two contrarotating vortices, whose position tends to remain stable (Fig. 11).

The comparison of the different flow field highlighted that, in this engine, the global flow structure is conserved even when one moves away from the engine head (on different measuring planes). Some differences in the flow orientation are evident at points 4 and 6. In fact, on the lowest plane, the velocity vectors move away from the y direction because of the greatest interaction with the flow entering the combustion chamber. There are negligible differences in the vectors' directions along diameter D_3. Moreover, the flow-field comparison shows that the velocity values increase with the distance from the engine head as the flow becomes more organized, while the flow symmetry tends slightly to reduce moving away from the head.

Then attention was focused on the velocity tangential components in order to correlate LDA measurements with a global swirl characterization carried out on the same engine head in a previous work. The data, obtained using an impulse swirl meter, show that the swirl coefficients increase with a rise in the valve lift and the throttle angle. More details are given in [24].

Figure 12 depicts the tangential velocity components and the corresponding standard deviations on the D_1 and D_3 diameters. The same configuration of the previous LDA campaign is adopted ($L_v/D_v = 0.178$, WOT, and three measuring planes).

The results display that the turbulence intensity distributes quite uniformly in spite of the uneven distribution of the mean flow. The outcome was in accordance with Kang and Reitz [30]. Typical rms values were

$$\sigma = 0.2 v_{\max}$$

where σ is the rms velocity and $v_{\max}$ is the largest tangential velocity.

Regarding the tangential velocity components, lower values on the diameter D_3 and larger velocities at the other measuring points were observed. Moreover, a marked symmetry was found on the diameter D_1 of the plane $H/B = 1/4$. Here a linear variation of the tangential velocity for each vortex was observed. Moving away from the engine head, as already observed for the velocity vectors, the velocities increased, as the flow field becomes more organized, while the flow symmetry was less accurate.

Furthermore, LDA measurements were used to calculate the angular momentum of the flux and the equivalent swirl coefficient applying Eqs. (8) and (9), respectively. To this purpose, in line with Kang and Reitz [30], a uniform axial flux was adopted to define the axial velocities. The hypothesis of uniform axial flux seems to be the most reasonable way to evaluate the angular momentum of the flux when, as in this work, a restricted number of measuring points are available and an uneven axial velocity distribution is present. In fact, the angular momentum of the flux depends strongly on the location of the measuring points and the risk is to alter its evaluation if the points are not well representative for all the regions in the cylinder. In this work the calculation is based on the 17 points located on the D_1 and D_3 diameters. The latter presents axial velocity components that are much larger than the corresponding values registered at the other locations (D_1, D_2,

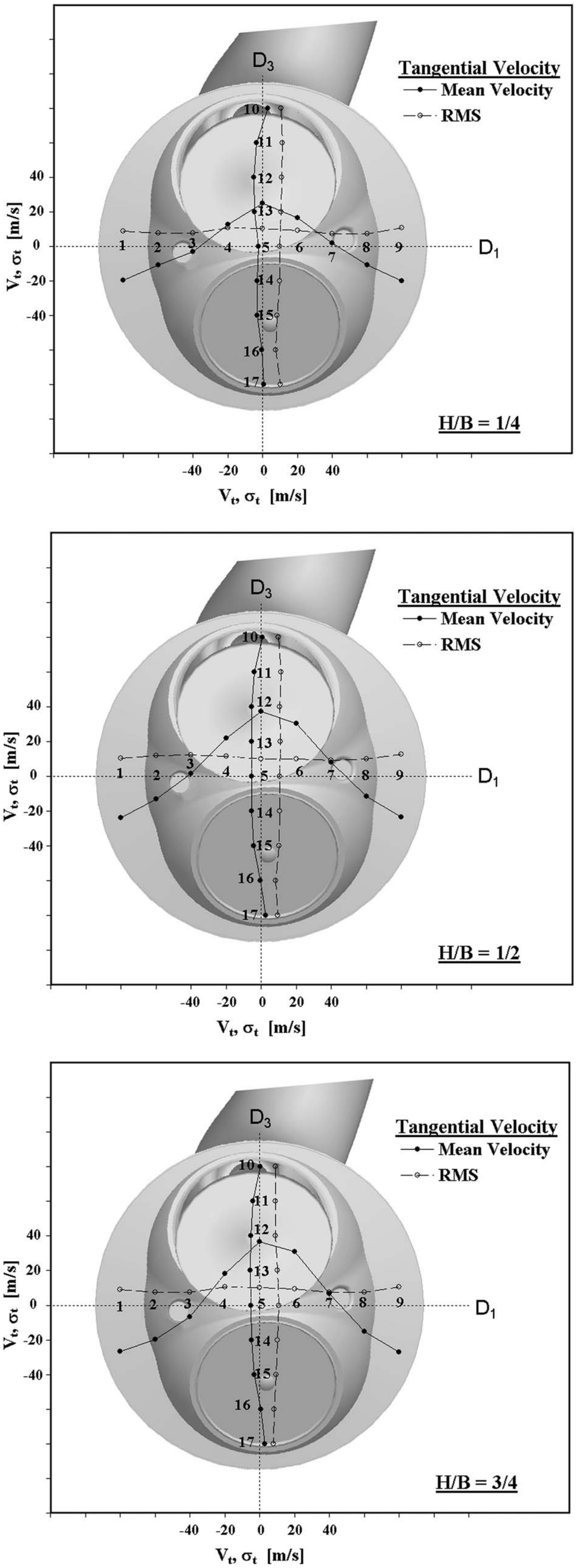


Fig. 12 LDA analysis: tangential and rms velocity on the three measuring planes

and D_4 diameters). The consequence is that, if the measured axial velocities are used, the calculation of the angular momentum is altered because the effect of the points distributed on the D_3 diameter is incorrectly amplified.

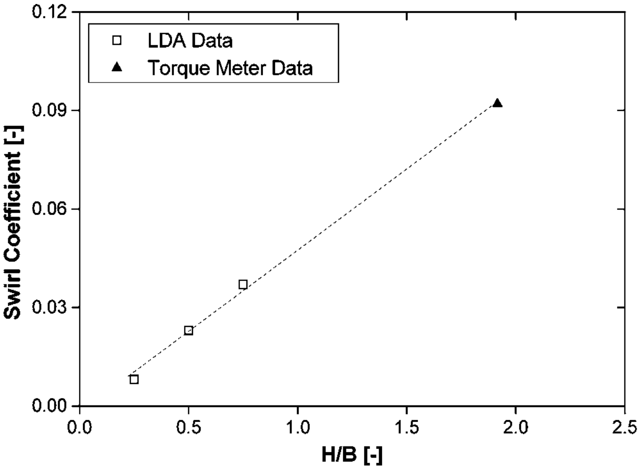


Fig. 13 LDA and swirl measurements correlation

Figure 13 shows the swirl coefficients obtained from LDA data and the value obtained by the swirl meter as a function of the distance between engine head and measuring plane. It is evident that the progressive slight reduction in the symmetry of the flow field determines an increase in the angular momentum and, as a consequence, in the swirl coefficient. In particular, the figure highlights that the value measured by the swirl impulse meter can be extrapolated with relatively good accuracy from the information obtained with the LDA technique. Figure 13 shows, in fact, that the LDA data and the torque meter data can be approximated by a straight line, even though it is difficult to give a physical explanation for this trend.

Conclusions

The fluid dynamic behavior of a high-performance four-stroke motorcycle engine was investigated at the steady flow rig. In particular, the intake phase was considered. Dimensionless flow coefficients and LDA measurements were used to define the global engine head breathability and to characterize the flow field inside the combustion chamber, respectively.

The global analysis showed that the throttle angle significantly influences the fluid dynamic efficiency of the engine head. As expected, the larger the throttle opening, the better the filling of the combustion chamber. Specifically, in the tested engine head for a throttle angle $F < 75$ deg, a continuous increase in the head breathability with the throttle opening was observed, while no further rises were registered for $F \geq 75$ deg. In addition, the influence of the throttling process up-surged with the valve lift. Furthermore, the flow coefficient analysis highlighted that a common trend in the fluid dynamic efficiency of the intake system exists, adopting an opportune new dimensionless plotting. In particular, all the experimental data were well approximated by a single curve (an exponential curve) independently from the analyzed configuration of the head. The extensibility of the proposed dimensionless rule was verified by analyzing a production four-cylinder automobile spark-ignition (SI) engine, and the same behavior was observed despite the high differences in the engines geometries and characteristics.

The LDA analysis showed the existence of two contrarotating vortices. The measurements carried out at different distances from the engine head depicted that the same flow structure was maintained for different measuring planes. Moreover, a progressive increase in the velocity values and a slight decrease in the symmetry of the two vortices was observed moving away from the head. Specifically, the progressive decrease in the flow symmetry was responsible for the increasing values of the angular momentum of the flux with the distance from the engine head. Then a correlation, in terms of dimensionless swirl coefficient, between

LDA measurements and a global swirl characterization carried out using an impulse swirl meter was proposed. The analysis proved that it is possible to extrapolate, with relatively good accuracy, global information from LDA data.

Acknowledgment

The authors would like to thank the Ducati Motor S.p.A. for the provision of the engine components and for the technical information.

Nomenclature

A_{ref} = reference area
B = cylinder bore
C_{abs} = absolute flow coefficient
C_f = flow coefficient
C_s = swirl coefficient
D_i = ith-measuring diameter
D_v = valve diameter
H = distance of the measuring plane from the head
L_v = valve lift
M = swirl momentum
T_0 = temperature pressure
a_0 = sound speed
$\dot{m}_{\text{meas}}$ = measured mass flow rate
$\dot{m}_r$ = reference mass flow rate
p_0 = ambient pressure
p_C = cylinder pressure
v_t = tangential velocity
v_a = axial velocity
v_0 = reference velocity
Δp = ambient-cylinder pressure drop
ϕ_{Ad} = dimensionless theoretical flow rate
ϕ_{Am} = dimensionless actual flow rate
λ = wavelength of the laser beam
ρ_0 = air density

References

[1] Heywood, J. B., 1998, *Internal Combustion Engine Fundamentals*, McGraw–Hill, New York.
[2] Zhijun, W., and Zhen, H., 2001, "In-Cylinder Swirl Formation Process In Four-Valve Diesel Engine," Exp. Fluids, **31**, pp. 467–473.
[3] Choi, W.-C., and Guezennec, Y. G., 1999, "Study of the Flow Field Development During the Intake Stroke in an IC Engine Using 2-D PIV and 3-D PTV," SAE Paper No. 1999-01-0957.
[4] Pignone, G. A., and Vercelli, U. R., 2003, *Motori ad Alta Potenza Specifica*, (in Italian), Giorgio Nada Editore, Milan.
[5] Algieri, A., and Bova, S., 2004, "Influence of Valve-Wall Distance on the Intake Flow in High Performance I.C.E.," SAE Paper No. 2004-01-1922, *Direct Fuel Injection, Engine Diagnostics, and New Developments in Powertrain Tribology, CVT, ATF & Fuel Economy*, SAE Special Publication No. SP-1891.
[6] Li, Y., Zhao, H., Leach, B., Ma, T., and Ladommatos, N., 2003, "Optimization of In-Cylinder Flow for Fuel Stratification in a Three-Valve Twin-Spark-Plug SI Engine," SAE Paper No. 2003-01-0653.
[7] Crnojevic, C., Decool, F., and Florent, P., 1999, "Swirl Measurements in a Motored Cylinder," Exp. Fluids, **26**, pp. 542–548.
[8] Arcoumanis, C., Nouri, J., Xu, H. M., and Stone, R., 1997, "Analysis of the Steady Flow Characteristics of Tumble-Generating Four-Valve Cylinder Heads," Opt. Diagnost. Eng., **2**(2), pp. 71–83.
[9] Jasak, H., Luo, J. Y., Kaludercic, B., Gosman, A. D., Echtle, H., Liang, Z., Wirbeleit, F., Wierse, M., Rips, S., Werner, A., Fernstrom, G., and Karlsson, H., 1999, "Rapid CFD Simulation of Internal Combustion Engines," SAE Paper No. 1999-01-1185.
[10] Bianchi, G. M., and Fontanesi, S., 2003, "On the Applications of Low-Reynolds Cubic k-ε Turbulence Models in 3D Simulations of ICE Intake Flows," SAE Paper No. 2003-01-0003.
[11] Laimbock, F. J., Klasnic, H., Grilc, S., Meister, G. F., and Dorfstatter, M. J., 1999, "SVL SDIS Mk. II: Low Cost Automotive FI Applied to 2-Stroke Engines for Future CARB-Regulations," SAE Paper No. 1999-01-3285.
[12] Xu, H., 2001, "Some Critical Technical Issues on the Steady Flow Testing of Cylinder Heads," SAE Paper No. 2001-01-13.
[13] Kang, K. Y., and Baek, J. H., 1995, "LDV Measurement and Analysis of Tumble Formation and Decay in a Four-Valve Engine," Exp. Therm. Fluid Sci., **11**, pp. 181–189.
[14] Blair, G. P., McBurney, D., McDonald, P., McKernan, P., and Fleck, R., 1998, "Some Fundamental Aspects of the Discharge Coefficients of Cylinder Porting and Ducting Restrictions," SAE Paper No. 980764.
[15] Bohac, S. V., and Landfahrer, K., 1999, "Effects of Pulsating Flow on Exhaust Port Flow Coefficients," SAE Paper No. 1999-01-0214.
[16] Blair, G. P., and Drouin, F. M. M., 1996, "Relationship Between Discharge Coefficients and Accuracy of Engine Simulation," SAE Paper No. 962527.
[17] Chan, V. S. S., and Turner, J. T., 2000, "Velocity Measurement Inside a Motored Internal Combustion Engine Using Three-Component Laser Doppler Anemometry," Opt. Laser Technol., **32**, pp. 557–566.
[18] Payri, F., Desantes, M., and Pastor, J. V., 1996, "LDV Measurements of the Flow Inside the Combustion Chamber of a 4-Valve D.I. Diesel Engine With Axisymmetric Piston-Bowls," Exp. Fluids, **22**, pp. 118–128.
[19] Bevan, K. E., and Ghandhi, J. B., 2004, "PIV Measurements of In-Cylinder Flow in a Four-Stroke Utility Engine and Correlation With Steady Flow Results," SAE Paper No. 2004-32-0005.
[20] El Tahry, S. H., Khalighi, B., and Kuziak, W. R. J., 1987, "Unsteady-Flow Velocity Measurements Around an Intake Valve of a Reciprocating Engine," SAE Paper No. 870593.
[21] Ohm, I. Y., and Cho, Y. S., 2000, "Mechanism of Axial Stratification and Its Effect in an SI Engine," SAE Paper No. 2000-01-2843.
[22] Stier, B., and Koochesfahani, M. M., 1999, "Molecular Tagging Velocimetry (MTV) measurements in Gas Phase Flows," Exp. Fluids, **26**, pp. 297–304.
[23] Hascher, H. G., Novak, M., Lee, K., Schock, H. J., Rezaei, H., and Koochesfahani, M., 1998, "An Evaluation of IC-Engine Flows With the Use of Modern In-Cylinder Measuring Techniques," 34th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit, July 13–15, Cleveland, OH.
[24] Carrera, M. A., 2004, "Caratterizzazione Fluidodinamica di Teste di MCI ad Elevate Prestazioni," Thesis (in Italian), Mechanics Department, University of Calabria.
[25] Amelio, M., Bova, S., and De Bartolo, C., 2000, "The Separation Between Turbulence and Mean Flow in ICE LDV Data: The Complementary Point-of-View of Different Investigation Tools," ASME J. Eng. Gas Turbines Power, **122**(4), pp. 579–587.
[26] Barbieri, G., and Zagami, S., 2000, "Progettazione e Realizzazione di un Banco di Flussaggio Stazionario per Rilievi LDA su Teste di M. C.I.," (in Italian), National ATI Congress.
[27] Algieri, A., Bova, S., and De Bartolo, C., 2005, "Experimental and Numerical Investigation on the Effects of the Seeding Properties on LDA Measurements," ASME J. Fluids Eng., **127**(3), pp. 514–522.
[28] Algieri, A., Amelio, M., and Bova, S., 2001, "Global and LDA Steady Flow Measurements in Two High-Performance I.C.E. Head Configurations," 5th International Conference on Internal Combustion Engines-ICE2001, SAE-NA Paper No. 2001-01-20.
[29] Auriemma, M., Caputo, G., Corcione, F. E., Valentino, G., and Riganti, G., 2003, "Fluid-Dynamic Analysis of the Intake System for a HDDI Diesel Engine by STAR-CD Code and LDA Technique," SAE Paper No. 2003-01-0002.
[30] Kang, K. Y., and Reitz, R. D., 1999, "The Effect of the Valve Alignment on Swirl Generation in a DI Diesel Engine," Exp. Therm. Fluid Sci., **20**, pp. 94–103.
[31] Weclas, M., Melling, A., and Durst, F., 1995, "Unsteady Intake Valve Gap Flow," SAE Paper No. 952477.
[32] Algieri, A., Bova, S., and De Bartolo, C., 2005, "Global Flow Characterization of a Four Cylinder Spark Ignition Engine During the Intake Phase," SAE Paper No. 2005-24-075.

Soot Formation Study in a Rapid Compression Machine

I. Kitsopanidis

W. K. Cheng

Sloan Automotive Laboratory,
Department of Mechanical Engineering,
Massachusetts Institute of Technology,
77 Massachusetts Avenue,
Cambridge, MA 02139

A rapid compression machine was used to study the soot formation process under diesel enginelike conditions. The apparatus creates accurately controlled conditions at the end of compression (uniform mixture, temperature, and well-defined mixture composition) and, by decoupling chemistry with mixing, provides an unambiguous data interpretation for kinetics study. The soot evolution was studied by the line-of-sight absorption method (at 632.8 nm*), which measured the soot volume concentration evolution in the initial stage of soot growth before the optical path became opaque. For a rich butane mixture at fuel equivalence ratio of 3, the ignition delay showed a negative temperature dependence at intermediate temperatures. The soot volume fraction showed an initial exponential growth, with a growth rate depending on the compressed charge fuel concentration. A substantial amount of soot was formed after the soot cloud became opaque. By weighing the total soot particles after the experiment, only* $\sim 10-15\ \%$ *of the soot mass was formed when the beam transmission was reduced to 5%. The final soot mass was* $\sim 15-18\ \%$ *of the total carbon mass for compressed charge density of* $250\ \text{mol/m}^3$ *and temperature from* 740 to 930 K. [DOI: 10.1115/1.2180279]

Introduction

Airborne particulate matter (PM) is a major air-quality concern because of its health effects and its impact on visibility [1]. As a result, the U.S. Environmental Protection Agency has established strict standards for the atmospheric PM levels [2]. A substantial source of PM in urban areas is vehicles, especially those powered by diesel engines. The PM emission from diesel engines originate from soot particles generated during the combustion process. These particles serve as condensation nuclei for low vapor pressure exhaust species, such as unburned or partially burned fuel and engine oil, where the particles grow further in mass [3]. Therefore, to effectively control PM emissions, it is important to understand the processes by which soot is formed in combustion.

Much of the current understanding of diesel combustion process was obtained from optical engine studies, such as those done at the Sandia National Laboratory [4]. Figure 1 shows a schematic of the combusting fuel jet. There are two major characteristics that are pertinent to the soot formation process:

1. *The fuel goes through a two-stage oxidation process*. The first stage is in the fuel-rich premixed zone just downstream of the liquid spray; then the products of this fuel-rich combustion are oxidized in the diffusion flame at the periphery of the plume.
2. *Soot is first formed in the fuel-rich premixed zone where fuel-air equivalence ratio* (Φ) *is in the range of* 2–4. Soot then grows in the environment of high-temperature fuel-rich combustion products by surface growth and agglomeration as it is convected toward the end of the plume.

Thus, soot formation is governed by the chemistry of the premixed fuel-rich mixture at Φ values of 2–4.

There are two classes of diesel soot formation studies:

(a) The observation of the process in a fuel jet that undergoes evaporation, mixing, ignition, and burning processes. The experiment could be done in an actual diesel engine, in a constant volume combustion vessel, in a burner, or in a rapid compression machine (RCM) [5–12]. Although the process is realistic in theses studies, it is difficult to extract quantitative kinetics information on soot formation because of the inherent nonuniformity and uncertainties in species concentrations and temperatures.

(b) The study of a reacting premixed mixture. The reaction is initiated by a shock wave in a shock tube, or by rapid compression in a rapid compression machine. In these studies, the operating environment is well controlled and quantitative data can be obtained for soot formation rates.

Data on soot formation rate and yield under premixed fuel-rich conditions have been obtained in shock tube studies at pressure and temperatures comparable to those found in diesel combustion [13–15]. Because of the low specific heat ratio of a fuel-rich hydrocarbon mixture, however, the shock tube studies were customarily done under heavily diluted conditions with a mixture consisting of 95–99.5 % argon so as to produce the correct thermal environment. As a result, the species concentrations were unrealistically low, and there was negligible heat release from the mixture in the process. To create a thermal and species concentrations environment that is representative of diesel combustion, rapid compression of a homogeneous fuel-rich mixture in an RCM is employed in this study.

The RCM is a single-stroke device which rapidly compresses a uniform mixture of fuel and air. The piston is locked at the end of the stroke to produce constant volume combustion. Figure 2 summarizes the temperature and fuel carbon concentration (in the combustible mixture) regimes of various experimental setups. It is apparent that the temperature and fuel carbon conditions achieved in an RCM are much more representative of those found in the soot formation process in a diesel engine compared to all other alternatives.

Soot detection in a combustion process can be achieved either with direct sampling or optical techniques. Molecular beam sampling combined with electron microscopy has been successful in identifying critical intermediates in burner types of experimental setups. Though sampling methods have been used directly in diesel engines [6], they are intrusive and it is difficult to obtain time-resolved data using the technique. The advancement of laser technology, on the other hand, made available nonintrusive tools for combustion diagnostics. The most common optical methods used for soot detection are the line-of-sight (LOS) absorption [7–11]

Contributed by the Internal Combustion Engines Division of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received July 11, 2005; final manuscript received October 24, 2005. Review conducted by M. Wooldbridge.

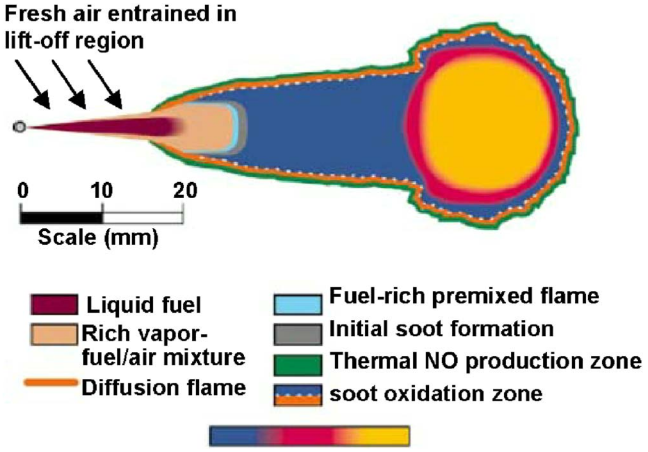


Fig. 1 Schematic of a diesel fuel spray illustrating the combustion processes [4]

and laser induced incandescence (LII) [12]. The latter method, however, could not provide the time resolution required for the single-shot RCM experiment. Therefore, LOS absorption was used to monitor the soot formation process.

The aim of the present work is to demonstrate the ability to study soot formation under diesel enginelike conditions created in an RCM using the LOS absorption technique. The RCM, which provides well-controlled conditions (pressure, temperature, and mixture composition), along with the nonintrusive technique of LOS absorption, are tools that could shed light on soot formation and provide valuable data for chemical kinetics modeling efforts.

Rapid Compression Machine

The RCM is a single-stroke device, which rapidly compresses a uniform mixture of fuel and oxidant to conditions similar to diesel engine operation. In contrast to an engine, the piston of the current RCM is locked in the compressed position to produce a constant volume environment. The mixture remains at these conditions for a longer period of time compared to a shock tube ($\sim$10 ms versus $\sim$1 ms); this time is only limited by heat losses. Thus, the thermal and composition conditions of the mixture under study are well defined.

The RCM, shown in Fig. 3, has a cylindrical combustion chamber with 5.08 cm bore and 1.27 cm clearance height. The piston starting position can be adjusted by changing the effective length of the hydraulic chamber with spacers (not shown in Fig. 3). The compression stroke then ranges from 15.24 cm to 20.32 cm, resulting in a compression ratio between 12.5 and 16.5. These values have taken into account the 3.1% crevice volume (compared to the clearance volume) in the piston and windows at the end of compression. For the data shown in this paper, the compression ratio was set to 16.5. The geometry and dimension of the current machine is comparable to others in the literature [16–18].

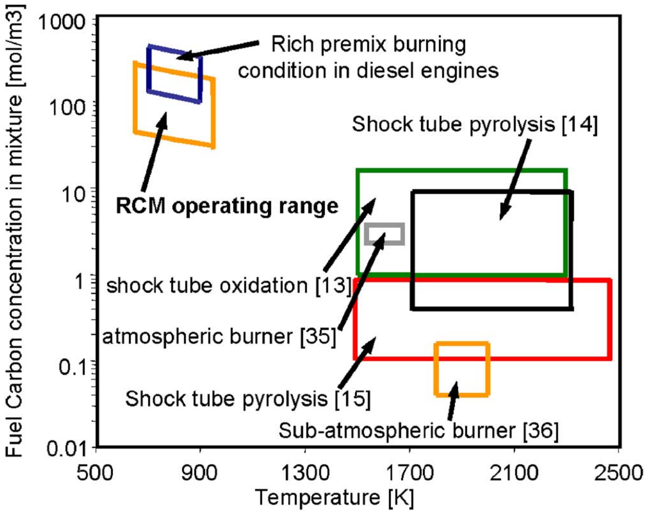


Fig. 2 Conditions for soot formation in diesel engines and in various experimental studies

The piston is pneumatically driven and uses a pin-and-groove mechanism in a hydraulic chamber to decelerate the piston at the end of the stroke [19]. In its retracted position, the pneumatic piston is pressurized (from the right of Fig. 3) to 17.5 bar gage pressure by a large tank of nitrogen, but it is locked in position by the highpressure oil (at 90 bar gage pressure) in the hydraulic chamber. When the oil pressure is released by a small solenoid valve, the piston moves forward, and thereby releases the locking hydraulic piston and starts the compression process.

During much of the stroke, the piston accelerates almost freely since the friction forces of the seals and oil are modest, and the pressure rise in the combustion chamber is not yet significant. The piston starts to decelerate significantly when the "pin" on the hydraulic piston enters the groove near the end of the stroke. The clearance between the pin and the groove is only 0.7 mm so that there is a large back pressure buildup that is responsible for stopping the piston with a reasonable cushion.

The calculated piston velocity and combustion chamber pressure (p) are shown in Fig. 4. The piston motion was obtained by using Newton's law; the substantial resistances to motion were the pressure buildup in the combustion chamber and the significant back pressure when the pin hit the groove. The latter value was obtained by a simple one-dimensional (1D) viscous flow calculation. The peak piston velocity was more than 30 m/s, and the total compression time was 15 ms. The effective compression time (de-

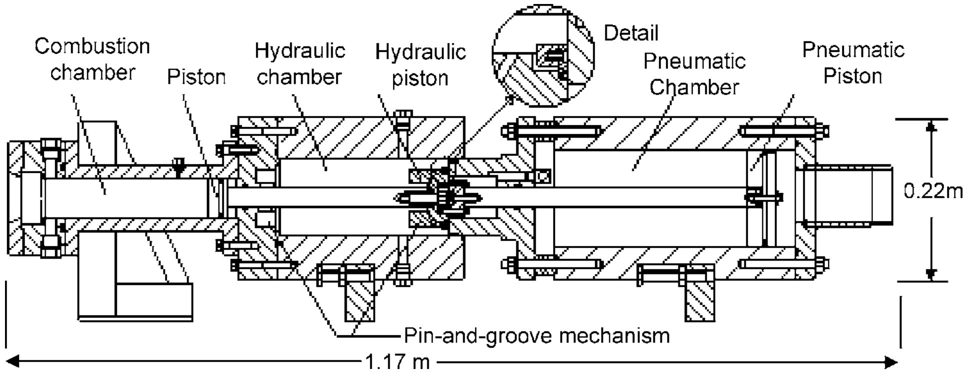


Fig. 3 Schematic of the rapid compression machine

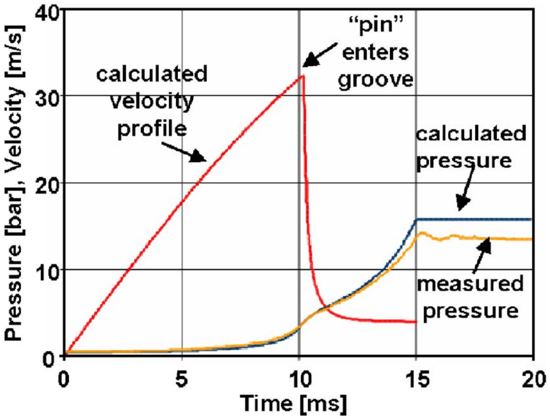


Fig. 4 Calculated piston velocity and comparison of calculated and observed pressure traces. *N*-butane/oxygen/argon mixture; $Ar/O_2=3.773$, $\Phi=3$; precompression pressure and temperature: 0.4 bar, 49°C.

fined as $p/(dp/dt)$ at the inflection point of the pressure trace) was 4 ms. The final compression time (defined as $p/(dp/dt)$ at the end of compression) was 0.4 ms. The calculated pressure (from adiabatic compression) agrees well with the experimental value except at the end of compression because of heat transfer. Because of the high deceleration (~3000 g) when the pin enters the groove, the finite impact velocity (~5 m/s) at the end of the stroke, and the large moving mass (6 kg), it is necessary to substantially anchor the RCM to prevent significant vibrations.

After reaching its final compression point, the piston was kept in place by the high-pressure driving gas so that constant volume combustion can be realized. For the driving pressure of 17.5 bar, the piston can hold a maximum combustion chamber pressure of 110 bar with the designed area ratio of the driving and the combustion chamber pistons.

The combustion chamber is equipped with four fused silica windows 1.27 cm dia around the circumference of the combustion chamber for line of sight measurements and a fully transparent cylinder head for imaging studies. Because the compression is rapid, the heat transfer to the wall is limited to a thin boundary layer so that there is a substantial adiabatic core in the charge. Then the compression temperature (T_{comp}) of this core may be obtained from the measured compression pressure (P_{comp}) and the initial condition (P_{ini} and T_{ini}) via the isentropic relation

$$\int_{T_{\text{ini}}}^{T_{\text{comp}}} c_P(T)\frac{dT}{T} = R\ln\left(\frac{P_{\text{comp}}}{P_{\text{ini}}}\right) \tag{1}$$

Here, $c_P(T)$ is the heat capacity per mole of the mixture at temperature T, and R is the universal gas constant. A Kistler 6125A piezoelectric pressure transducer was used to measure the combustion chamber pressure at sampling frequency of 100 kHz.

The mixture was prepared manometrically. The combustion chamber was evacuated before the gaseous mixture components were introduced. A Baratron 622A pressure transducer with accuracy of 0.1 Torr was used to meter each mixture component. The combustion chamber, connecting lines and fuel reservoirs were all insulated, heated, and temperature controlled to accommodate nonvolatile fuels, which required an initial temperature to be higher than the highest dew point corresponding to the partial pressure of the respective component in the mixture.

To obtain high compression temperatures, an argon/oxygen mixture was used instead of air as the oxidant. The Ar/O_2 molar ratio was fixed at 3.773, which is the same as the N_2/O_2 ratio in air. Then, at the end of compression, a relatively wide range of conditions similar to the range found in diesel engines (10–30 bar, 700–900 K) was accomplished by varying the initial temperature and pressure before compression commenced.

After each run, the soot in the combustion chamber was collected and weighed, although there might still be trace amount of soot left on the O-rings and crevices. Then the combustion chamber, window, and piston surfaces were cleaned with acetone. Special care was taken at the crevices where soot could accumulate. Then all O-rings were cleaned with acetone, dried, and lubricated with vacuum grease before reassembly. The piston ring was cleaned only after every 20 runs because the procedure involved substantial disassembly of the apparatus. After the apparatus was reassembled, the combustion chamber was heated to the desired temperature and evacuated before the new mixture was introduced. Because of the high vapor pressure of acetone (normal boiling point 56.6°C), there should not be any substantial solvent materials left after the system was pumped down.

Laser Extinction Measurement of Soot Volume Concentration

Real-time line-of-sight (LOS) absorption using a helium-neon laser beam at a wavelength (λ) of 632.8 nm is used to determine the soot volume concentration evolution. The method is well established [20,21] and briefly summarized as follows.

Since the soot particle sizes at the nucleation and initial growth stages are less than, or of the order of, 50 nm [21–23], which is much less than λ/π, the interaction of the particles and the light beam is in the Rayleigh regime; the extinction of the laser beam is predominantly due to absorption rather than scattering. Assuming that the particles are spherical with diameter D, the absorption cross section is proportional to D so that the extinction coefficient is proportional to D^3 and, hence, proportional to the soot particle volume. Then the attenuation of the laser beam may be related to the LOS integral of the soot volume fraction

$$\ln\left(\frac{I}{I_0}\right) = \frac{6\pi}{\lambda}\,\text{Im}\left(\frac{m^2-1}{m^2+2}\right)\int_0^L f_v dx \tag{2}$$

In Eq. (2), I_0 and I are the incident and transmitted laser beam intensities, m is the complex refractive index of the soot particles, L is the path length, and the soot volume concentration f_v is defined by

$$f_v = n\int_0^\infty \frac{\pi D^3}{6}p(D)dD \tag{3}$$

where n is the soot particle number per unit volume and $p(D)$ is the probability distribution of the particle diameter. If the soot cloud is uniform in space, the value of f_v may be obtained from the measurement of I/I_0 as

$$f_v = \frac{\lambda\ln(I/I_0)}{6\pi L\,\text{Im}\left(\frac{m^2-1}{m^2+2}\right)} \tag{4}$$

It is assumed that the refractive index m is known and remains constant during the course of soot evolution. The refractive index of soot has only been determined in the bulk. Chang and Charalampopoulos [24] measured the complex refractive index based on dynamic light scattering at a wavelength of 632.8 nm to be $1.8-0.58i$ at 10 mm above the burner surface and $1.62-0.47i$ at 6 mm. Lee and Tien [25] calculated the complex refractive index based on the multivariable dispersion model to be $1.9-0.55i$ at 632.8 nm without finding any significant temperature effect. The maximum difference in the calculated soot volume concentration using any of the above values is ±6%. To be consistent with most recent studies, the value suggested by Chang and Charalampopoulos at 10 mm above the burner was used in this study.

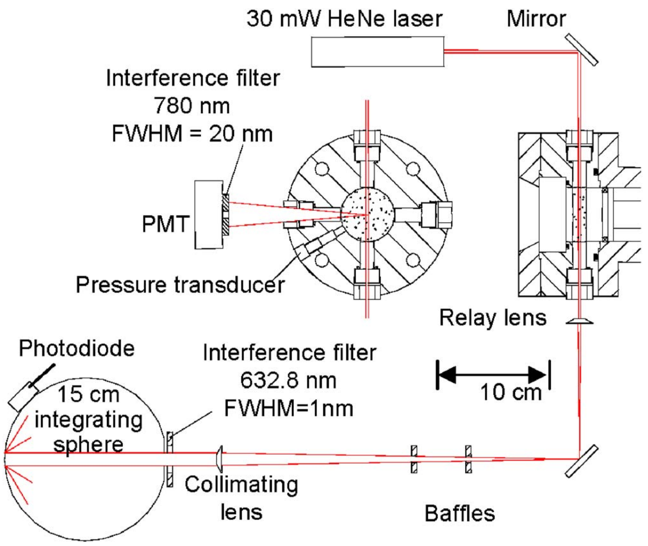


Fig. 5 Schematics of optical setup

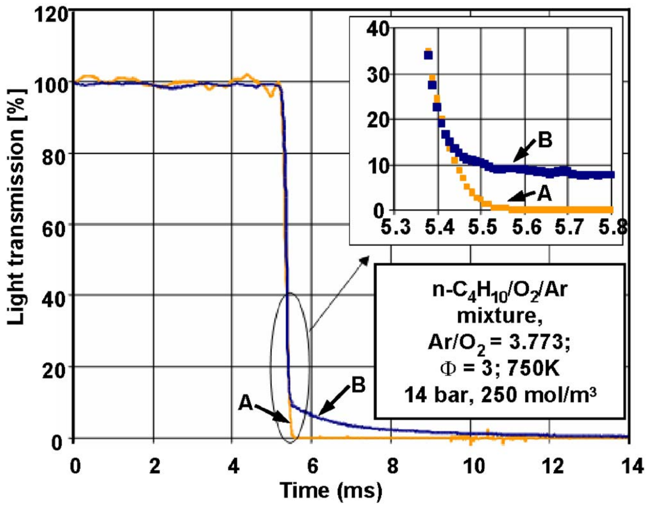


Fig. 6 Effect of soot luminosity on apparent transmission: (*a*) signal with optical arrangement of Fig. 5 and (*b*) with detector and filter at 5 cm from combustion chamber window

Optical Setup. The schematic of the optical setup is shown in Fig. 5. A 30 mW He–Ne laser beam passes through the combustion chamber, where it is attenuated by fuel-rich combustion generated soot particles. The attenuated beam is then connected to the photodiode detector at a rather large distance (~0.6 m) by relay lenses, baffles, and spectral filter.

The above optical arrangement is necessary because there is a significant amount of combustion-generated light. During combustion, light emission takes place due to chemiluminescence and soot luminosity. The former arises from radicals decaying back to equilibrium energy levels after being raised to an excited energetic state. In hydrocarbon oxidation, such species are OH (310 nm), C_2 (516 nm), CH_2O (368–470 nm), and especially CH (390, 430 nm) [26]. Dec and Espey, after conducting experiments in a diesel engine, found that chemiluminescence peaks at around 430 nm and is almost zero above 620 nm, with HCO speculated as the only possible species responsible at higher wavelengths [27]. By using a laser light at 632.8 nm, the chemiluminescense interference can be eliminated. Soot luminosity, on the other hand, arises from soot particles, which act as radiating gray bodies after being generated. The emission of this light has a broadband spectrum, and for typical rich combustion temperatures of 1800–2000 K, the emissive power can be significant at 632.8 nm. The level is comparable to that of the laser light during the later stage of combustion since the latter is strongly attenuated by the soot particles.

To minimize soot luminosity detected by the photodiode, a narrow band (FWHM of 1 nm) interference filter was used and the detector was moved away from the test section, since the soot luminosity contribution drops with the square of distance. The signal improvement is shown in Fig. 6, in which the detected signals under the same compression conditions are compared. Curve A shows the transmission signal with the optical setup of Fig. 5; curve B shows the one obtained with photodiode and interference filter at ~5 cm from the combustion chamber window. The significantly higher apparent transmission in curve B is due to the contribution from soot emissions. Experiments were also run with the optical arrangement of Fig. 5 with the laser beam off. No luminosity was detected; thus, the arrangement was sufficient to reduce the soot luminosity contribution to a negligible level.

With a long optical path, significant noise is induced at the detector due to beam wander. The beam motion is caused by mechanical vibrations of the optical train and by beam steering due to the refractive index gradients inside the combustion chamber. To alleviate these effects, an integrating sphere is used to collect the laser light so that the detector signal is not sensitive to beam wander. Note also that since the interference filter is partial to the angle of incidence, it has to be placed in front of the integrating sphere; a collimating lens is further used to direct the whole beam at normal incidence to the filter.

Confirmation of Absorption Signal due to Soot Particles. There are reports in the literature that query whether light absorption at 632.8 nm is evidence of soot particles presence. Microscopy studies suggest that there are large molecules other than soot particles that may absorb light at this wavelength [28]. Other studies suggested that very young soot particles may not absorb light at this wavelength [29,30]. Though the distinction between soot precursors (which are large molecules) and soot particles is arbitrary; there does not seem to be any gaseous species (i.e., molecules that are not large enough to exhibit macroscopic properties, such as black-body radiation) that absorb light at 632.8 nm [31]. Then absorption at this wavelength may be interpreted as an indication of soot presence. To confirm this point, a photomultiplier tube (PMT) was placed above the top window of the RCM with an interference filter at 780 nm (FWHM (full width at half maximum) =10 nm), as shown in Fig. 5. Laser light scattering is blocked by the filter, and, as already explained earlier, there is no chemiluminescent at this wavelength. Thus, the only light that can be detected by this PMT comes from particle gray-body radiation, which indicates the presence of a "particle." Figure 7 shows the PMT detected signal and the light transmission simultaneously. The luminosity signal starts at the same time as light absorption. This fact confirms that the absorbing media is indeed in the particulate phase.

Results

In the following, typical results from the RCM are used to demonstrate the performance of the machine and the type of data it provides. Comprehensive data sets and comparison of data with models of ignition delay and soot formation kinetics will be presented in later publications.

Figure 8 shows the pressure trace and light transmission for five consecutive tests of a typical fuel-rich mixture. All tests were conducted with *n*-butane/O_2/Ar mixture, at Ar/O_2=3.773 and Φ=3. The conditions at the end of compression were the same for all tests: P_{comp}=14 bar, T_{comp}=765 K, and total molar concentration n_{comp}=250 mol/m^3. After a total compression time of ~15 ms, there is a period of preignition reaction, during which the rate of heat release cannot keep pace with the heat losses and

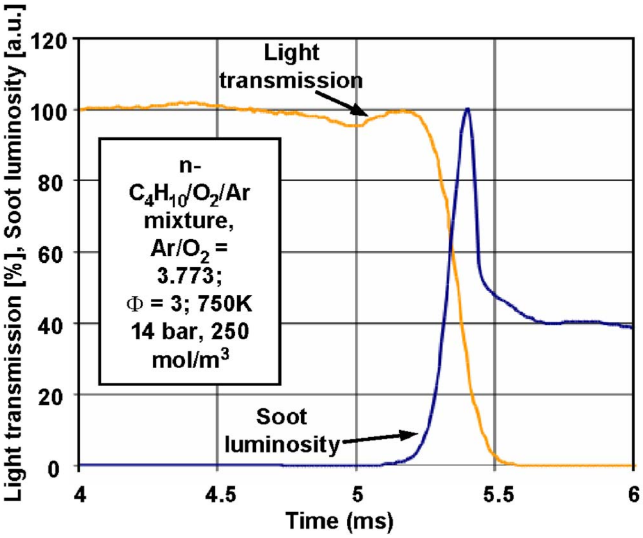


Fig. 7 Soot luminosity and light transmission

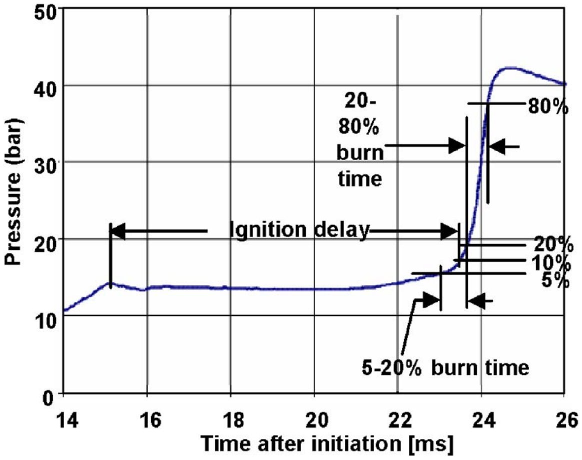


Fig. 9 Ignition delay and combustion times. *N*-butane/oxygen/argon mixture; Ar/O$_2$=3.773, Φ=3; compressed conditions: 765 K, 14 bar, 250 mol/m^3.

the pressure drops slightly before the mixture ignites at ~8 ms after the end of compression. The heat release is in two stages, with a slower rate first followed by a rapid one. These observations are consistent with what is reported in the literature for most hydrocarbon fuels [32]. The pressure drop from the peak value is due to heat transfer to the cold walls when the combustion is over. Note that the repeatability is excellent for both the pressure trace and the light transmission; for example, the variability of the ignition delay is ±0.15 ms at 95% confidence level.

Ignition Delay and Combustion Time. The ignition delay and combustion time may be derived from the pressure curve as illustrated in Fig. 9. The ignition delay is defined as the time from the end of compression to the time at the 10% heat release point. To assess the combustion rate, the 5–20% and the 20–80% burn duration are chosen to reflect the early and later parts of combustion. The threshold levels used in these definitions are somewhat arbitrary but are chosen to give well-defined results. (For example, because of the uncertainty in the heat release curve, the 80% heat release point is much better defined than the 90% one.) Although the main purpose of the experimental study is on soot formation, the ignition delay, and the combustion times are essential for validating the chemical kinetics of the soot formation process. These values are also archival since there is not very much fundamental data on ignition delay and combustion times for hydrocarbon fuels under fuel-rich conditions.

The ignition delay data for *n*-butane at fuel equivalence ratio (Φ) of 3 and a compression molar density of 250 mol/m^3 are shown in Fig. 10. Note that to vary the compression temperature, the set of experiments was done at different initial temperatures; the initial pressures were adjusted to produce the same compression molar density. The Arrhenius plot is typical of data from hydrocarbons: at low temperatures, the ignition delay decreases rapidly with increase of temperature. At intermediate temperatures, the delay stays fairly constant or even increases slightly because of the well-known negative temperature coefficient (NTC) regime of the chemistry involved. At higher temperatures, the delay decreases again with increase in temperature. The repeatability of the ignition delay value is ±2%.

The 5–20% and 20–80% burn times are plotted in Fig. 11. The early burn time followed an Arrhenius temperature dependence with activation energy of 3.38 kcal/mol. Both values are of the order of a fraction of to 1 ms. The repeatability of the burn times is ±12%.

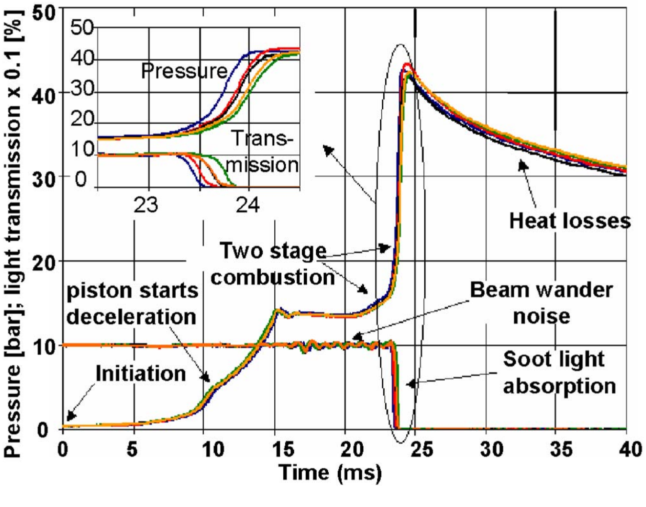


Fig. 8 Repeatability of data; five repeats. *N*-butane/oxygen/argon mixture; Ar/O$_2$=3.773, Φ=3; compressed conditions: 750 K, 14 bar, 250 mol/m^3.

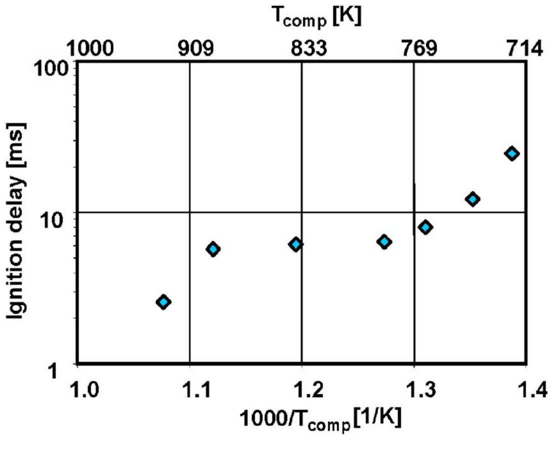


Fig. 10 Ignition delay as function of compression temperature. *N*-butane/oxygen/argon mixture; Ar/O$_2$=3.773, Φ=3; compressed molar density: 250 mol/m^3.

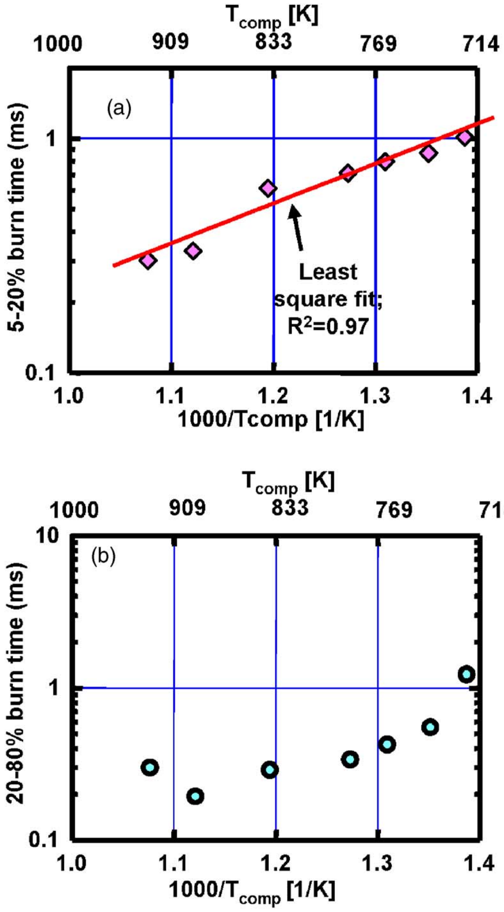


Fig. 11 Combustion time as a function of compression temperature: (*a*) 5–20 % burn time and (*b*) 20–80 % burn time. *N*-butane/oxygen/argon mixture; $Ar/O_2=3.773$, $\Phi=3$; compressed molar density: 250 mol/m^3.

Soot Evolution. The instantaneous soot volume concentration f_v may be calculated the transmission signal according to Eq. (4). Because of the beam wander noise (see Fig. 8) and that the signal-to-noise ratio was not favorable at high beam attenuation, the signal was only deem valid when the transmission was between 90% and 5%. It should be noted that the 5% transmission point does not mark the end of soot formation. Since the mixtures are sufficiently rich, the combustion chamber can become almost opaque well before the soot formation process ends. Collecting and weighing the soot after the end of the experiment and converting the mass to an equivalent final soot volume concentration (using a nominal density of 1.8 g/cm^3), it was found that, depending on the operating conditions, the value of f_v at the 5% transmission point corresponds to only 10–15 % of the final value. The absorption experiments, therefore, only provide information about the initial stage of soot formation. This information, however, is critical for validating the soot formation kinetics.

The time evolutions of the soot volume concentration are shown in Fig. 12, in which the value of f_v is normalized by the molar concentration of carbon atoms at end of compression (denoted by $[C]_{\text{comp}}$) to obtain a soot yield (SY). Specifically, the definition of SY is

$$SY = \frac{\rho_s f_v}{W_C[C]_{\text{comp}}} \tag{5}$$

where $\rho_s=1.8$ g/cm^3 is the presumed density of soot, $W_C=12$ g/mol is the atomic weight of carbon, and $[C]_{\text{comp}}$ is the molar carbon concentration of the charge at end of compression. The time zero corresponded to the point of 90% transmission, while the end point corresponded to 5% transmission. For the data in case A (at $\Phi=3$, compression temperature and density of 765 K and 250 mol/m^3), SY grew exponentially in time. For case B (same condition as case A, except that the compressed density was lower, at 190 mol/m^3), the exponential growth was observed for the initial part of the trace. The growth rate was then slower at a later time. The initial exponential growth was observed in all the test cases with Φ ranging from 2.5 to 4.5 over a large range of compression temperatures and densities.

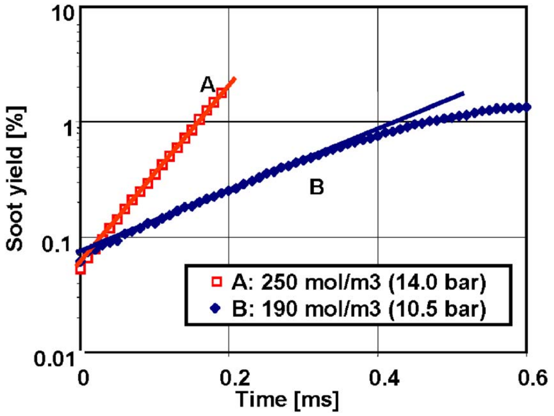


Fig. 12 Soot formation history at two different gas densities. *N*-butane/oxygen/argon mixture; $Ar/O_2=3.773$, $\Phi=3$; compressed temperature: 765 K. Time zero corresponded to point of 90% light transmission.

The empirically observed exponential growth of SY cannot be simply explained. If it is assumed that in this period, the particles are predominantly in the growth stage (i.e., nucleation is complete and agglomeration is not yet important so that the particle number density (n) is approximately constant), then the growth of soot particle is governed by the condensation of the gaseous hydrocarbon (HC) species, such as polycyclic aromatic hydrocarbons (PAH) and C_2H_2, derived from the fuel molecules. These species have a higher C/H ratio than the parent molecule; they act as building materials to the surface growth of the young soot particles [33]. For a spherical soot particle of diameter D that is smaller than the mean free path, the growth in volume should then be proportional to the collision area

$$\frac{d}{dt}\left(\frac{\rho \pi D^3}{6}\right) = \frac{(\pi D^2)}{4}[\text{HC}]W_{\text{HC}}v_{\text{HC}} \tag{6}$$

where ρ is the density of the soot particle, [HC] and W_{HC} are the molar concentration and molecular weight of the condensing soot precursors, and v_{HC} is the relative velocity between the precursor molecules and the soot. If the values of ρ, [HC], W_{HC}, and v_{HC} are constant, the soot volume concentration f_v, which is proportional to nD^3, should, according to Eq. (6), increase as t^3, which is much slower than exponential.[1]

Since the soot particles are known to be porous, one may make the tacit assumption that the density decreases as the particle grows, where $\rho \propto 1/D$. Then D grows exponentially in time and the soot yield (SY) is:

$$SY \propto \exp\{(aW_{\text{HC}}v_{\text{HC}})[\text{HC}]t\} \tag{7}$$

where a is a proportional constant. Thus, the growth rate would be proportional to the precursor concentrations $[HC]$. Equation (7)

---

[1] At a charge density of ~200 mol/m^3, the mean free path is ~10 nm. The large particles may approach the continuum limit. Then the growth rate is proportional to $t^{3/2}$, which is even slower.

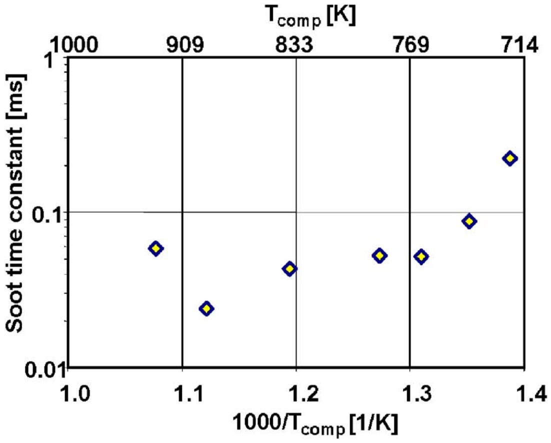


Fig. 13 Soot formation time constant as a function of compression temperature. *N*-butane/oxygen/argon mixture; $Ar/O_2 = 3.773$, $\Phi = 3$; compressed molar density: 250 mol/m^3.

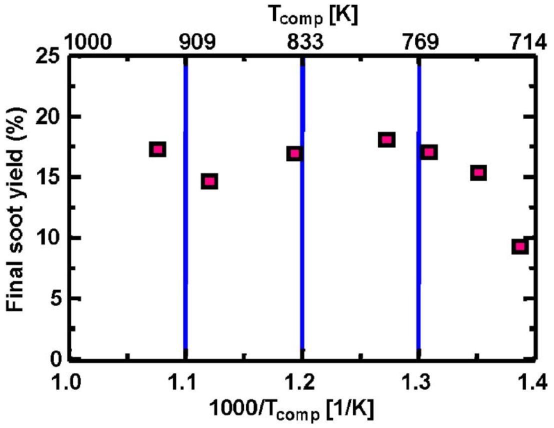


Fig. 14 Final soot yield as a function of temperature. *N*-butane/oxygen/argon mixture; $Ar/O_2 = 3.773$, $\Phi = 3$; compressed molar density: 250 mol/m^3.

explains qualitatively the features in Fig. 12:

(i) The growth is exponential in time,
(ii) The slope of the plotted lines increases with charge density increase, which supplies a higher value of [HC].
(iii) For Case B, which has a lower charge density, the growth rate at a later time is slower than exponential because of the depletion of [HC]. (Similar behavior was not observed in case A because the optical path had become opaque before the slow down in the growth rate.)

Quantitative, however, it is much more difficult to reach agreement. We expect that [HC] should be proportional to the amount of fuel present, so the growth rate should be proportional to the fuel concentration (which, for fixed Φ value, scales with the compressed charge density). Experimentally, the ratio of the two slopes in Fig. 12 is 2.90; this value disagrees with the ratio of the charge density at end of compression: ratio$=250/190=1.32$. Thus, the dependence on the fuel concentration is complex and there is no simple closure. Quantitative comparisons could only be sought through detailed modeling of the formation and oxidation of both the precursors and soot [34].

Since the observed initial growth of the soot volume concentration is exponential, a soot formation time constant (equal to the inverse of the logarithmic soot yield growth rate) may be obtained. The time constant is shown in Fig. 13 as a function of compression temperature at $\Phi = 3$ and compression density of 250 mol/m^3. The values are below 0.1 ms at compression temperatures above 740 K.

Final Soot Yield. As mentioned before, the soot cloud becomes opaque as time progresses so that the LOS absorption measurement could not provide information about the later part of the soot evolution process. The total final soot mass, however, could be obtained by carefully collecting all the soot particles from the inside of the apparatus after the experiment and weighing them with a microbalance (Ohaus Model EP214 with resolution of 0.1 mg). Depending on the experimental condition, the collected soot mass was of the order of 5–30 mg. In spite of the crudeness of the method, repeatability was excellent; the uncertainty was ±5% (based on five repeats). The final soot yield (total final soot mass normal by total mass of carbon atoms in the combustion chamber) for a mixture at $\Phi = 3$, and compressed charge density of 250 mol/m^3 is plotted in Fig. 14 as a function of temperature. Except for the experiment at the lowest temperature (722 K), the final soot yield at these conditions is not sensitive to temperature (740–930 K); the value is ~15–18 %.

Conclusions

The rapid compression machine (RCM) was demonstrated to be an appropriate apparatus for studying soot formation under fuel-rich conditions in an environment similar to that in actual diesel engines. Unlike engines studies, however, the RCM provides a uniform charge with well-defined temperature and pressure so that the results can be interpreted quantitatively. Unlike shock tube studies, the fuel/air mixture is not diluted so that soot formation process can be studied under realistic levels of fuel carbon concentrations. The soot volume concentration evolution was observed in real time by line-of-sight absorption measurement at 632.8 nm. Luminosity from the soot radiation interfered with the measurement; the interference was alleviated by placing the detector far from the test section and the use of baffles and a narrowband pass-interference filter. Data were reported from compression ignition of a *n*-butane/oxygen/argon mixture with argon to oxygen molar ratio of 3.773 (equal to that of nitrogen to oxygen ratio in air) under fuel-rich condition at a fuel equivalence ratio of 3. The ignition delay versus temperature data showed a negative temperature coefficient region at intermediate temperatures. The combustion was in two stages: a slower one followed by a rapid one.

For all the experiments, the instantaneous volume concentration grew exponentially in time, initially. When the fuel concentration of the compressed charge was sufficiently low, slower growth rate was observed in a later stage before the soot cloud became opaque. There was, however, no simple model that would explain the initial exponential growth behavior. The time constant associated with this growth was fast—less than 0.1 ms for temperatures above 740 K at fuel equivalence ratio of 3 and charge density of 250 mol/m^3. There was significant further soot formation after the laser beam was blocked by the soot cloud. The soot concentration at 5% transmission was ~10–15 % of the final soot concentration obtained by weighing the soot particles collected from the combustion chamber after the experiment. The final soot mass was ~15–18 % of the carbon mass in the charge.

Acknowledgment

This work was supported by the Center of Airborne Organics of the U.S. Environmental Protection Agency, and an Industrial Consortium for Engine and Fuels Research. The members of the Consortium were DaimlerChrysler, Delphi, ExxonMobil, Ford Motor, and GM. I. Kitsoponidis was also supported by a fellowship from Cummins Engine Co.

References

[1] U.S. Environmental Protection Agency, 1966, “EPA’s Proposal for the Particulate Matter Standard: Fact Sheet.”

[2] Anon, 1997, “National Ambient Air Quality Standards for Particulate Matter; Final Rule,” Federal Register, 62, p. 38652.

[3] Heywood, J. B., 1988, *Internal Combustion Engine Fundamentals*, McGraw-Hill, New York.

[4] Dec, J., 1997, “A Conceptual Model of DI Diesel Combustion Based on Laser-Sheet Imaging,” SAE Paper No. 970873.

[5] Senda, J., Choi, D., Iwamuro, M., Fujimoto, H., and Asai, G., 2002, “Experimental Analysis on Soot Formation Process in DI Diesel Combustion Chamber by Use of Optical Diagnostics,” SAE Paper No. 2002-01-0893.

[6] Pungs, A., Pischinger, S., Backer, H., and Lepperhoff, G., 2000, “Analysis of the Particle Size Distribution in the Cylinder of a Common-Rail DI Diesel Engine During Combustion and Expansion,” SAE Paper No. 2000-01-1999.

[7] Tree, D., and Dec, J., 2001, “Extinction Measurements of In-Cylinder Soot Deposition in a Heavy Duty DI Diesel Engine,” SAE Paper No. 2001-01-1296.

[8] Song, K. H., Lee, Y., and Litziger, A., 2000, “Effects of Emulsified Fuels on Soot Evolution in an Optically Accessible DI Diesel Engine,” SAE Paper No. 2000-01-2794.

[9] Hentschel, W., and Richter, J. U., “Time Resolved Analysis of Soot Formation and Oxidation in a Direct-Injection Diesel Engine for Different EGR Rates by an Extinction Method,” SAE Paper No. 952517, 1995.

[10] Wiartalla, A., Backer, H., and Durnholz, M., “Influence of Injection System Parameters on Spray Development, Combustion and Soot Formation by Optical Measurement Techniques in a Model Combustion Chamber,” SAE Paper No. 950233, 1995.

[11] Miysmoto, N., Ogawa, H., Goto, N., and Sasaki, H., “Analysis of Diesel Soot Formation Under Varied Ignition Lag With a Laser Light Extinction Method,” SAE Paper No. 900640, 1990.

[12] Dec., J., and Espey, C., “Ignition and Early Soot Formation in a DI Diesel Engine Using Multiple 2D Imaging Diagnostics,” SAE Paper No. 950456, 1995.

[13] Kellerer, H., Koch, R., and Wittig, S., 2000, “Measurements of the Growth and Coagulation of Soot Particles in a High-Pressure Shock Tube,” Combust. Flame, **120**, pp. 188–199.

[14] Bauerle, S., Karasevich, Y., Slavov, S., Tanke, D., Tappe, M., Thienel, T., and Wagner, H. G., 1994, “Soot Formation at Elevated Pressures and Carbon Concentrations in Hydrocarbon Pyrolysis,” *25th Symposium (International) on Combustion*, The Combustion Institute, Pittsburgh, pp. 627–634.

[15] Parker, T. E., Foutter, R. R., and Rawlins, W. T., 1990, “Soot Initiation and Particle Growth in the Pyrolysis of Toluene at High Inert Gas Pressures,” AIP Conf. Proc., **208**, pp. 481–486.

[16] Griffiths, J. F., Jiao, Q., Kordylewski, W., Schreiber, M., Meyer, J., and Knoche, K. F., 1993, “Experimental and Numerical Studies of Ditertiary Butyl Peroxide Combustion at High Pressures in a Rapid Compression Machine,” Combust. Flame, **93**, pp. 303–315.

[17] Carlier, M., Corre, C., Minetti, R., Pauwels, J. F., Ribaucour, M., and Sochet, L. R., 1990, “Autoignition of Butane: A Burner and a Rapid Compression Machine Study,” *23 Symposium (International) on Combustion*, The Combustion Institute, Pittsburgh, pp. 1753–1758.

[18] Park, P., and Keck, J. C., 1990, “Rapid Compression Machine Measurements of Ignition Delays for Primary Reference Fuels,” SAE Paper No. 900027.

[19] Affleck, W. S., and Thomas, A., 1968, “An Opposed Piston Rapid Compression Machine for Preflame Reaction Studies,” Proc. Inst. Mech. Eng., **183**, pp. 365–381.

[20] Jones, A. R., 1999, “Light Scattering for Particle Characteristics,” Prog. Energy Combust. Sci., **25**, pp. 1–53.

[21] Haynes, B. S., and Wagner, H. G., 1981, “Soot Formation,” Prog. Energy Combust. Sci., **7**, 229–273.

[22] Graham, S. C., 1976, “The Collisional Growth of Soot Particles at High Temperatures,” *16th Symposium (International) on Combustion*, The Combustion Institute, Pittsburgh, pp. 663–669.

[23] Amann, C. A., and Siegla, D. C., 1982, “Diesel Particulates—What They Are and Why?,” Aerosol Sci. Technol., **1**, pp. 73–101.

[24] Chang, H., and Charalampopoulos, T. T., 1990, “Determination of the Wavelength Dependence of Refractive Indices of Flame Soot,” Proc. R. Soc. London, Ser. A, **430**, pp. 577–591.

[25] Lee, S. C., and Tien, C. L., 1981, “Optical Constants of Soot in Hydrocarbon Flames,” *18th Symposium (International) on Combustion*, The Combustion Institute, Pittsburgh, pp. 1159–1166.

[26] Gaydon, A. G., 1974, *The Spectroscopy of Flames*, 2nd ed., Chapman and Hall, London.

[27] Dec, J., and Espey, C., 1998, “Chemiluminescence Imaging of Autoignition in a DI Diesel Engine,” SAE Paper No. 982685.

[28] Wersborg, B. L., Fox, L. K., and Howard, J. B., 1975, “Soot Concentration and Absorption Coefficient in a Low-Pressure Flame,” Combust. Flame, **24**, pp. 1–10.

[29] Vaglieco, B. M., Merola, S. S., D’Anna, A., and D’Alessio, A., 2002, “Spectroscopic Analysis and Modeling of Particulate Formation in a Diesel Engine,” J. Quant. Spectrosc. Radiat. Transf., **73**, pp. 443–450.

[30] D’Anna, A., Violi, A., D’Alessio, A., and Sarofim, A. F., 2001, “A Reaction Pathway for Nanoparticle Formation in Rich Premixed Flames,” Combust. Flame, **127**, pp. 1995–2003.

[31] Bjorseth, A., 1985, *Handbook of Polycyclic Aromatic Hydrocarbons*, Marcel Dekker, New York.

[32] Glassman, I., 1996, *Combustion*, 3rd ed., Academic Press, New York.

[33] Richter, H., Howard, J. B., 2000, “Formation of Polycyclic Aromatic Hydrocarbons and Their Growth to Soot—A Review of Chemical Reaction Pathways,” Prog. Energy Combust. Sci., **26**, pp. 565–608.

[34] Kitsopanidis, I., 2004, “Experimental and Computational Study of Soot Formation Under Diesel Engine Conditions,” Ph.D. thesis, Department of Mechanical Engineering, MIT, July, available at http://libraries.mit.edu/

[35] Harris, S. J., and Weiner, A. M., 1983, “Determination of the Rate Constant for Soot Surface Growth,” Combust. Sci. Technol., **32**, pp. 267–275.

[36] Wersborg, B. L., and Howard, J. B., 1975, “Soot Concentration and Absorption Coefficient in a Low-Pressure Flame,” Combust. Flame, **24**, pp. 1–10.