

Transactions of the ASME®

Editor
LEE S. LANGSTON (2006)
Assistant to the Editor
LIZ LANGSTON
Associate Editors
Fuels and Combustion Technologies
S. GOLLAHALLI (2004)
Internal Combustion Engines
D. ASSANIS (2005)
Nuclear
R. DUFFY (2004)
International Gas Turbine Institute
IGTI Review Chair
R. NATOLE (2001)
E. BENVENUTI (2002)
H. R. SIMMONS (2003)
Combustion and Fuels
P. MALTE (2003)
Structures and Dynamics
N. ARAKERE (2004)
M. MIGNOLET (2005)

BOARD ON COMMUNICATIONS
Chair and Vice-President
OZDEN OCHOA

OFFICERS OF THE ASME
President, **REGINALD VACHON**
Executive Director, **V. R. CARTER**
Treasurer, **R. E. NICKELL**

PUBLISHING STAFF
Managing Director, Engineering
THOMAS G. LOUGHLIN
Director, Technical Publishing
PHILIP DI VIETRO
Managing Editor, Technical Publishing
CYNTHIA B. CLARK
Manager, Journals
JOAN MERANZE
Production Coordinator
JUDITH SIERANT
Production Assistant
MARISOL ANDINO

Transactions of the ASME, Journal of Engineering
for Gas Turbines and Power (ISSN 0742-4795) is published
quarterly (Jan., April, July, Oct.) by The American
Society of Mechanical Engineers, Three Park Avenue, New
York, NY 10016. Periodicals postage paid at New
York, NY and additional mailing offices. POSTMASTER:
Send address changes to Transactions of the ASME, Journal
of Engineering for Gas Turbines and Power, c/o THE
AMERICAN SOCIETY OF MECHANICAL ENGINEERS, 22
Law Drive, Box 2300, Fairfield, NJ 07007-2300.

CHANGES OF ADDRESS must be received at Society
headquarters seven weeks before they are to be effective.
Please send old label and new address.

STATEMENT from By-Laws. The Society shall not be
responsible for statements or opinions advanced in papers
or ... printed in its publications (B7.1, par. 3).

COPYRIGHT © 2003 by the American Society of Mechanical
Engineers. For authorization to photocopy material for
internal or personal use under circumstances not falling
within the fair use provisions of the Copyright Act,
contact the Copyright Clearance Center (CCC), 222
Rosewood Drive, Danvers, MA 01923, Tel: 978-750-8400,
www.copyright.com.

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

Journal of Engineering for Gas Turbines and Power

Published Quarterly by The American Society of Mechanical Engineers

VOLUME 125 • NUMBER 3 • JULY 2003

TECHNICAL PAPERS

Gas Turbines: Controls, Diagnostics & Instrumentation

- 613** Controller Design for a Gas Turbine Using Periodic Output Feedback
A. K. Chakrabarti and B. Bandyopadhyay
- 617** Application of Fuzzy Logic for Fault Isolation of Jet Engines
(2001-GT-0013)
R. Ganguli
- 624** Interpretation of Weighted-Least-Squares Gas Path Analysis Results
(2002-GT-30025)
D. L. Doel
- 634** Setting Up of a Probabilistic Neural Network for Sensor Fault Detection
Including Operation With Component Faults (2002-GT-30030)
C. Romesis and K. Mathioudakis
- 642** Nonlinear Engine Component Fault Diagnosis From a Limited Number of
Measurements Using a Combinatorial Approach (2002-GT-30031)
N. Aretakis, K. Mathioudakis, and A. Stamatis

Gas Turbines: Cycle Innovations

- 651** Degradation Effects on Combined Cycle Power Plant Performance—Part
I: Gas Turbine Cycle Component Degradation Effects (2001-GT-0388)
A. Zwebek and P. Pilidis
- 658** Degradation Effects on Combined Cycle Power Plant Performance—Part
II: Steam Turbine Cycle Component Degradation Effects (2001-GT-0389)
A. Zwebek and P. Pilidis

Gas Turbines: Coal, Biomass, Alternative Fuels and Combustion & Fuels

- 664** Revisiting the Inhibition of Vanadium-Induced Hot Corrosion in Gas
Turbines (2001-GT-0005)
E. Rocca, P. Steinmetz, and M. Moliere

Gas Turbines: Combustion and Fuels

- 670** Flashback on Lean Prevaporized Premixed Combustion: Nonswirling
Turbulent Pipe Flow Study (2001-GT-0053)
O. Schäfer, R. Koch, and S. Wittig
- 677** Time Domain Simulation of Combustion Instabilities in Annular
Combustors (2002-GT-30063)
C. Pankiewicz and T. Sattelmayer
- 686** On-Line Fuel Deoxygenation for Coke Suppression (2002-GT-30071)
L. J. Spadaccini and H. Huang
- 693** Flashback Limits for Combustion-Induced Vortex Breakdown in a Swirl
Burner (2002-GT-30075)
M. Kröner, J. Fritz, and T. Sattelmayer
- 701** Impact of Ethane and Propane Variation in Natural Gas on the
Performance of a Model Gas Turbine Combustor (2002-GT-30080)
R. M. Flores, V. G. McDonell, and G. S. Samuelsen

(Contents continued on inside back cover)

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

- 709 High Power Density Silicon Combustion Systems for Micro Gas Turbine Engines (2002-GT-30082)
C. M. Spadaccini, A. Mehra, J. Lee, X. Zhang, S. Lukachko, and I. A. Waitz
- 720 Two-Dimensional Flow and NO_x Emissions in Deflagrative Internal Combustion Wave Rotor Configurations (2002-GT-30085)
K. Pekkan and M. R. Nalim

Gas Turbines: Heat Transfer

- 734 The Two-Flux Model Applied to Radiative Transfer and Forced Convection in the Laminar Boundary Layer on a Flat Plate
F. Malpica, N. Moreno, and A. Tremante
- 742 Gas Turbine Engine Durability Impacts of High Fuel-Air Ratio Combustors—Part 1: Potential for Secondary Combustion of Partially Reacted Fuel (2002-GT-30077)
S. P. Lukachko, D. R. Kirk, and I. A. Waitz
- 751 Gas Turbine Engine Durability Impacts of High Fuel-Air Ratio Combustors—Part II: Near-Wall Reaction Effects on Film-Cooled Heat Transfer (2002-GT-30182)
D. R. Kirk, G. R. Guenette, S. P. Lukachko, and I. A. Waitz

Gas Turbines: Manufacturing, Materials & Metallurgy

- 760 Assessment of Current Turbine Engine High Cycle Fatigue Test Methods
K. L. Nichol

Gas Turbines: Structures and Dynamics

- 766 Reducing Lateral Vibration of a Rotor Passing Through Critical Speeds by Phase Modulating
S.-M. Wang, Q.-S. Lu, and E. H. Twizell
- 772 A Reduced-Order Model for Evaluating the Effect of Rotational Speed on the Natural Frequencies and Mode Shapes of Blades (2000-GT-0611)
P. Marugabandhu and J. H. Griffin

Gas Turbines: Vehicular & Small Turbomachines

- 777 A Gas Turbine-Based Hybrid Vehicle—Part II: Technological and Configuration Issues
R. Capata, E. Ciofarelli, and E. Sciubba

Internal Combustion Engines

- 783 Effect of Fuel Injection Timing Relative to Ignition Timing on the Natural-Gas Direct-Injection Combustion
Z. Huang, S. Shiga, T. Ueda, H. Nakamura, T. Ishima, T. Obokata, M. Tsue, and M. Kono
- 791 Statistical Correlation Between the Crankshaft's Speed Variation and Engine Performance—Part I: Theoretical Model
D. Taraza
- 797 Statistical Correlation Between the Crankshaft's Speed Variation and Engine Performance—Part II: Detection of Deficient Cylinders and Mean Indicated Pressure Calculation
D. Taraza
- 804 Numerical Study of Nonreacting Gas Turbine Combustor Swirl Flow Using Reynolds Stress Model
S. L. Yang, Y. K. Siow, B. D. Peschke, and R. R. Tacina
- 812 Air Fuel Ratio Estimation Using In-Cylinder Pressure Frequency Analysis
N. Cavina and F. Ponti
- 820 Wear Assessment in a Biodiesel Fueled Compression Ignition Engine
A. K. Agarwal, J. Bijwe, and L. M. Das
- 827 Combustion in Gas Fueled Compression: Ignition Engines of the Dual Fuel Type
G. A. Karim
- 837 Homogeneous Charge Compression Ignition Operation With Natural Gas: Fuel Composition Implications
J. Hiltner, R. Agama, F. Mauss, B. Johansson, and M. Christensen
- 845 Endothermic Phenomena in Diesel Particulate Filters During Regeneration
N. Kyriakis, T. Manikas, P. Pistikopoulos, and E. Vouitsis

TECHNICAL BRIEFS

- 851 Subsonic Elector Nozzle Limiting Flow Conditions
L. J. De Chant
- 854 Computing CHEMKIN Sensitivities Using Complex Variables
N. Butuk and J.-P. Pemba

A. K. Chakrabarti
B. Bandyopadhyay
e-mail: bijnan@ee.iitb.ernet.in

Systems and Control Engineering, IIT Bombay,
Mumbai 400076, India

Controller Design for a Gas Turbine Using Periodic Output Feedback

The advent of electronic controllers in the field of gas turbine control has allowed the implementation of sophisticated control algorithms. This paper proposes one such method, to design a controller using piecewise constant periodic output feedback, which has multifarious advantages, and because the states are not required for feedback, complete pole-assignability is guaranteed and the method is easily implementable.
[DOI: 10.1115/1.1581891]

1 Introduction

Feedback control has been an essential part of gas turbine control from the very beginning. However, it is only recently that computers have been used to implement engine controls. Many engines even today use hydromechanical controllers which implement the desired control strategy in terms of cams and mechanical integrators. The changeover to electronic controllers allowed the implementation of improved control strategies, which would be difficult to implement mechanically.

Various control techniques like state feedback control, linear quadratic control, etc., have been proposed by numerous authors including Watts [1] and De Hoff [2]. The problem with these control techniques is that the state feedback controller requires the availability of the entire state vector or needs an estimator, which may not be feasible in the case of gas turbines.

On the other hand, static output feedback requires only the measurement of system output. It is one of the most investigated problems in control theory and application since it represents the simplest control that can be realized in practice. Secondly, many problems involving synthesizing dynamic controllers can be formulated as static output feedback problems involving augmented plants. However, complete pole-assignment and guaranteed closed-loop stability has still not been obtained using static output feedback.

The periodic output feedback problem, which has been considered in this paper, has the features of static output feedback and also guarantees complete pole-assignability. Since the gains in periodic output feedback are piecewise constant the method is easily implementable, and compared to static output feedback, a much larger class of systems can be stabilized.

The linear model data required for most of the control applications was available for an F100 turbofan engine, from the work by De Hoff [2]. However, to arrive at this linear representation, one has to go through the necessary steps of simulation, followed by the generation of the linear models at the selected flight points. These aspects have been dealt with in brief in the first few sections.

Gas turbines in general are higher-order systems, and the cost and complexity involved in controller design is also high. In this paper a well-known result is utilized to reduce the system order by retaining the dominant eigenvalues and then arriving at a state feedback gain, which stabilizes the reduced system. This, when aggregated and applied to the higher order system guarantees closed-loop stability, which is an advantage of this approach. A

simultaneously stabilizing periodic output feedback gain sequence realizes the output injection gains obtained for the various linear models.

The concept of periodic output feedback, the procedure followed to arrive at the robust periodic output feedback gain K , and the results have been presented in the later sections.

2 Nonlinear Simulation

The nonlinear digital simulation of the F100 engine includes both steady-state component descriptions as well as dynamic elements to represent gas volume, thermal capacitance, and rotor inertial effects. The steady-state component operating characteristics (maps) are tied together by dynamic or time-varying relationships based on the conservation of mass and energy. Point-to-point calculations are made, using the rotor inertial effects, enclosed volume capacitive effects, and transient heat transfer effects. The rotor dynamics, pressure dynamics, and transient heat transfer effects account for the 16 states present in the model.

The steady-state maps and dynamic relationships combine to form a set of simultaneous, nonlinear differential equations. E.g., rotor inertial effects lead to the dynamic equation

$$\Delta Q_T - \Delta Q_C = J \frac{dN}{dt} \quad (1)$$

or

$$W_T \Delta h_T - W_C \Delta h_C = J \frac{dN}{dt} \quad (2)$$

Since simulation involves assumption of certain values in the beginning, it is not necessary that the above equation will be satisfied. In that case the simulation program generates an error variable.

$$E_1 = W_C \Delta h_C + J \frac{dN}{dt} - W_T \Delta h_T, \quad (3)$$

which can be solved digitally using recursive techniques to arrive at the correct values. A detailed description of the simulation techniques has been discussed by Seller [3].

3 Linear Model Generation and Selection

Most of the analytical design approaches used by control designers require linear descriptions of steady-state and transient characteristics of the engine. Linear approximations to nonlinear systems can be obtained at a particular operating point (a particular equilibrium condition of power setting and flight condition) for small excursions about that point.

A number of altitude/Mach number flight points have to be selected, with emphasis on extreme conditions of pressure, tem-

Contributed by the Controls, Diagnostics, and Instrumentation Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the CDI Division Nov. 2000; final revision received by the ASME Headquarters Dec. 2002. IGTI Committee Chair: G. Sheard.

perature, and at locations where control system limitations exist. Variations in power settings at these flight points account for a larger number of linear models, as shown by Hackney [4].

4 Brief Introduction to Periodic Output Feedback

The approach we use here is based on the result by Chammas and Leonides [5] and further work carried out by Werner and Furuta [6]. The poles of a discrete system can be arbitrarily assigned (with the natural restriction that they appear in conjugate pairs) using a piecewise constant periodic output feedback. The idea is to sample the output of the system at time instants $t = k\tau$, $k = 0, 1, \dots$. The control signal is generated according to

$$u(t) = K_l y(k\tau), k\tau + l\Delta \leq k\tau + (l+1)\Delta, K_{l+N} = K_l \quad (4)$$

for $l = 0, 1, \dots$, where a sampling interval τ is divided into N subintervals, $\Delta = \tau/N$.

Let the system sampled at the rate $1/\tau$, be $\{\Phi_\tau, \Gamma_\tau, C\}$ and that the system sampled at the rate $1/\Delta$, be $\{\Phi, \Gamma, C\}$.

Now the closed-loop system becomes

$$x(k\tau + \tau) = (\Phi^N + \Gamma K C)x(k\tau), \quad (5)$$

where

$$\Gamma = [\Phi^N \Gamma \dots \dots \dots \Gamma], \quad (6)$$

and

$$K^T = [K_0^T \dots \dots \dots K_{N-1}^T]. \quad (7)$$

The sufficient condition for the existence of the gain matrix K , such that the closed-loop system has any arbitrary set of poles, is that we choose N to be greater than or equal to the controllability index ν of the system. The controllability index ν is the minimum γ for which the matrix $[\Gamma, \Phi\Gamma, \dots, \Phi^{\gamma-1}\Gamma]$ has full rank.

The problem has now taken the form of a static output feedback problem. Let G be an output injection gain matrix for the system $\{\Phi_\tau, \Gamma_\tau, C\}$, such that

$$\rho(\Phi^N + GC) < 1, \quad (8)$$

where $\rho(\cdot)$ denotes the spectral radius. By observability one can choose an output injection gain G to achieve any desired self-conjugate set of eigenvalues for the closed-loop matrix $(\Phi^N + GC)$, and from $N \geq \nu$ it follows that one can find a periodic output feedback gain, which realizes the output injection gain G , by solving

$$\Gamma K = G \quad (9)$$

for K .

The controller obtained from Eq. (9) will give the desired behavior, but might require excessive control action. To reduce this effect we relax the condition that K exactly satisfies the above linear equation and include a constraint on the gain K . Thus we arrive at the equation

$$\|\Gamma K - G\| < \rho. \quad (10)$$

This can be formulated in the framework of linear matrix inequalities as follows:

$$\begin{bmatrix} -\rho^2 I & (\Gamma K - G) \\ (\Gamma K - G)^T & -I \end{bmatrix} < 0. \quad (11)$$

The periodic output feedback controller obtained by the above method requires only constant gains and hence is easier to implement.

5 Procedure Followed

The procedure undertaken for designing the controller is as follows.

5.1 Description of Plants and Control Law. Six linear models were available at various flight points/power settings. The

models comprise of 16 states. Two single-input single-output cases are considered. In the first case, the input taken is the "main burner fuel flow," which is the natural control parameter for any gas turbine. The output in this case is the "compressor speed." In the second case, the input is the same as the previous case, while the output is the "turbine inlet temperature." Both the outputs considered are critical parameters for the operation of the gas generator (core) portion of the F100 turbofan engine. The linear models available at the six operating points were found to be completely controllable and observable.

5.2 Model Reduction. Numerous model reduction techniques have been discussed in [7]. However, the approach we have used here is based on the method proposed by Lamba [8].

Consider a linear continuous time-invariant system, represented by

$$\begin{aligned} \dot{x} &= Ax + Bu, \\ y &= Cx \end{aligned} \quad (12)$$

where $x \in R^n$, $u \in R$, $y \in R$, and A , B , C are of appropriate dimensions.

If we sample this system at τ instants, we get the discrete representation

$$\begin{aligned} x(k+1) &= \Phi_\tau x(k) + \Gamma_\tau u(k), \\ y(k) &= Cx(k). \end{aligned} \quad (13)$$

The adjoint or the dual for the above system would be

$$\begin{aligned} \hat{x}(k+1) &= \Phi_\tau^T \hat{x}(k) + C^T \hat{u}(k), \\ \hat{y}(k) &= \Gamma_\tau^T \hat{x}(k). \end{aligned} \quad (14)$$

Assuming that the eigenvalues of Φ_τ^T are real and distinct, one can arrive at the model form of the above system using a suitable transformation matrix M . On truncation of the transformed system, we get the reduced r th-order modal as

$$\begin{aligned} \hat{z}_r(k+1) &= \Phi_1 \hat{z}_r(k) + C_1 \hat{u}(k), \\ \hat{y}_r(k) &= \Gamma_1 \hat{z}_r(k). \end{aligned} \quad (15)$$

5.3 Output Injection and Periodic Feedback Gains. Using the `dlqr` command in MATLAB, one can get a state feedback gain S_r , which makes the reduced closed-loop system $(\Phi_1 + C_1 S_r)$ stable. On aggregation, it can be shown that the higher-order closed-loop system $(\Phi_\tau^T + C^T S)$ is also stable, where $S = S_r [I_r : 0_{r \times (n-r)}] M^{-1}$. But the eigenvalues of $(\Phi_\tau^T + C^T S)$ and $(\Phi_\tau^T + C^T S)^T$ are the same. So $(\Phi_\tau + S^T C)$ will also be stable.

Thus $S^T \equiv G$ is the output injection gain for the system in Eq. (13). Then, $\Gamma K = G$, can be solved for K , using LMI framework.

In this case six LMIs are solved for the equal number of plants to get a common K , which would stabilize and give a good response for all the plants.

6 Simulation Results

The open-loop impulse response of the plants indicate that the system is inherently stable with the eigenvalues lying inside the unit circle of the Z -plane.

As discussed in the earlier section, the periodic output feedback gain K is arrived at, using the output injection gain G . For low values of τ , it is observed that the gain K is very high. However, by manipulation of the Q and R weighing matrices in the performance index chosen for the reduced model, one can get suitable values of K with good closed-loop response.

The dominant time constant for the plants, is found to vary in the range of 0.1 to 1.67. A robust K , in terms of stability and performance, has been obtained for $\tau = 0.13$ in the first case and for $\tau = 0.24$ (higher than the dominant time constants of two plants) in the second case. The gain values are low and realizable.

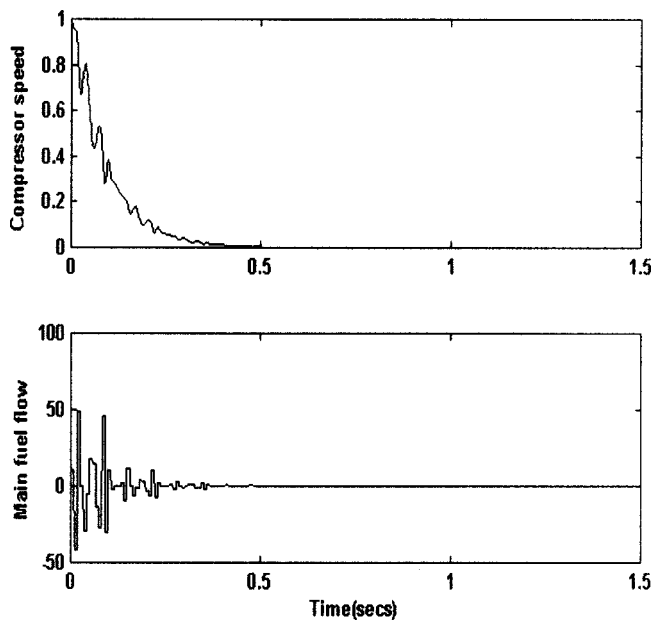


Fig. 1 Closed-loop response and control effort in case of Plant 2

The gain sequences obtained are

Case I— $K^T = [10.90 \ -42.34 \ 50.84 \ -1.40 \ -30.35 \ -3.25 \ 18.57 \ 12.70 \ -14.03 \ -24.33 \ 42.93 \ -28.73 \ 10.36 \ -2.10 \ 0.22 \ -0.009]$ and

Case II— $K^T = [-0.95 \ 4.65 \ -8.34 \ 4.88 \ 3.66 \ -5.56 \ -0.79 \ 4.93 \ -2.66 \ -0.67 \ 1.43 \ -0.74 \ 0.20 \ -0.03 \ 0.0022 \ -0.000061]$.

Sample closed-loop response of the plants, along with the control efforts, are shown in Figs. 1 to 4.

Figure 1 indicates the closed-loop response and the control effort in case of Plant 2, with the output being the “compressor speed” and the input being the “main burner fuel flow.” The response is stable and settles early. However, the initial fluctuation in the control effort is slightly high. Since the settling time of the control effort is very low, it is practically acceptable.

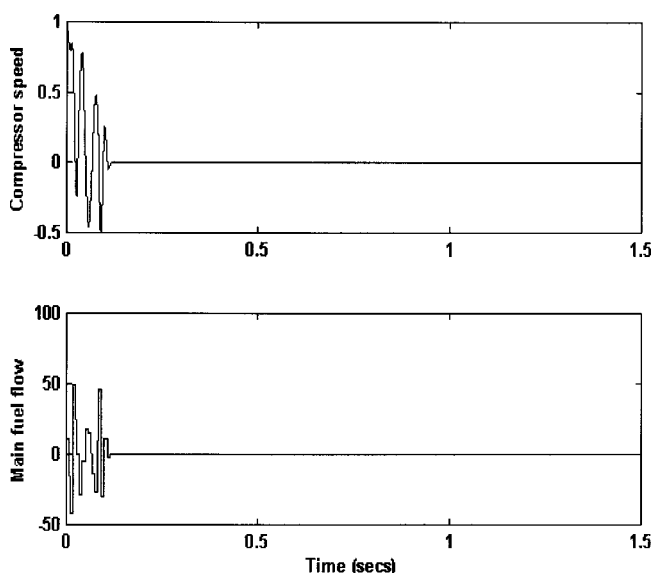


Fig. 2 Closed-loop response and control effort in case of Plant 5

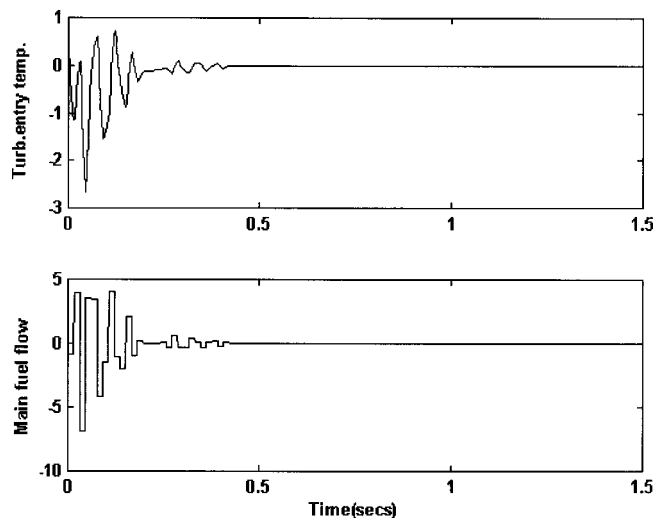


Fig. 3 Closed-loop response and control effort in case of Plant 2

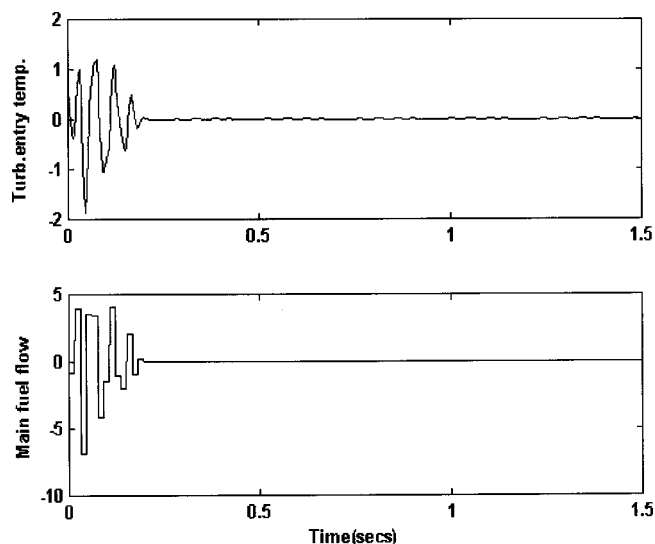


Fig. 4 Closed-loop response and control effort in case of Plant 5

Figure 2 indicates the closed-loop response and the control effort in case of Plant 5, with the output and the input same as in Fig. 1. The results are similar to Fig. 1, except that the response and the control effort settling time is further reduced.

Figure 3 indicates the closed-loop response and control effort in case of Plant 2, with the output being the “turbine inlet temperature.” The response is stable and settles early. The control effort also settles early and has low fluctuations.

Figure 4 indicates the closed-loop response and the control effort in case of Plant 5, with the output and input same as in Fig. 3. the results are similar as in Fig. 3.

7 Conclusion

The controller design for an F100 turbofan engine using piecewise constant periodic output feedback has been discussed in this paper, with a brief mention of the simulation and linear model generation and selection aspects.

In practice, not all of the states are available for measurement. In this case, the optimal control requires to design the state observer. This technique increases the implementation cost and reduces the reliability of the control system. Hence, it is desirable to

go for a periodic output feedback design, which has the advantages of static output feedback and also guarantees complete pole-assignability. With the added features of ease of implementation and the ability to stabilize a larger class of systems, this technique appears to be a viable method for the design of a multivariable controller for gas turbines.

The periodic output feedback gain K , arrived at via reduced order models, is observed to perform satisfactorily.

A = area or system matrix in continuous domain
 B, C = state-space matrices in continuous domain
 E = error variable
 G, S = output injection gain matrices
 J = polar moment of inertia
 K = periodic output feedback gain matrix
 N = speed or subinterval of τ
 Q = torque
 $\dot{W}$ = mass flow
 h = specific enthalpy
 u = input
 x = state
 y = output
 Φ, Γ, C = state-space matrices in discrete domain
 ρ = scalar
 ν = controllability index
 $\rho(.)$ = spectral radius

Superscripts

T = transpose
 $\wedge$ = dual system

Subscripts

r = reduced-order system
 τ = system discretized at τ instants

References

- [1] Watts, J. W., Dwan, T. E., and Brockus, C. G., 1992, "Optimal State-Space Control of a Gas Turbine Engine," *ASME J. Eng. Gas Turbines Power*, **114**, pp. 763–767.
- [2] De Hoff, R. L., Hall, W. E., Adams, R. J., and Gupta, N. K., 1977, "F 100 Multivariable Control Synthesis," *Systems Control Inc., AFAPL-TR-77-35*, Vol. I and II.
- [3] Sellar, J. F., and Danielle, C. J., 1975, "DYNGEN-A Program for Calculating Steady-State and Transient Performance of Turbo-Jet and Turbo-Fan Engines," Lewis Research Center, NASA.
- [4] Hackney, R. D., and Miller, R. J., 1977, "Engine Criteria and Models for Multivariable Control System Design," AFAPL, OH.
- [5] Chammas, A. B., and Leondes, C. T., 1979, "Pole-Assignment by Piecewise Constant Output Feedback," *Int. J. Control*, **29**(1), pp. 31–38.
- [6] Werner, H., and Furuta, K., 1995, "Simultaneous Stabilization Based on Output Measurement," *Kybernetika*, **31**(4), pp. 395–411.
- [7] Mahmoud, M. S., and Singh, M. G., 1981, *Large Scale Systems Modeling*, Pergamon Press, London.
- [8] Lamba, S. S., and Rao, S. V., 1974, "On Suboptimal Control via the Simplified Model of Davison," *IEEE Trans. Autom. Control*, **AC-19**, pp. 448–450.

Application of Fuzzy Logic for Fault Isolation of Jet Engines

R. Ganguli

Assistant Professor,
Department of Aerospace Engineering,
Indian Institute of Science,
Bangalore 560 012, India

A fuzzy logic system is developed for gas turbine module fault isolation. Inputs to the fuzzy logic system are measurement deviations of gas path parameters from a "good" baseline engine. The gas path measurements used are exhaust gas temperature, low and high rotor speed, and fuel flow. These sensor measurements are available on most jet engines. The fuzzy logic system uses rules developed from performance influence coefficients to isolate the module fault while accounting for uncertainty in the gas path measurements. Tests with simulated data show the fuzzy system isolates module faults with accuracy of over 95%. In addition, the fuzzy logic system shows good performance even with poor quality data. Additional pressure and temperature measurements between the compressor and before the burner help to increase the accuracy of fault isolation at high levels of uncertainty and when modeling assumptions weaken. [DOI: 10.1115/1.1470481]

Introduction

Gas turbine performance diagnostics involves the accurate detection, isolation, and estimation of engine module performance, engine system problems, and instrumentation problems using measurements from the engine gas path. Discernable changes in gas path measurements from a baseline "good engine" are used to obtain changes in engine performance from the baseline state. Historically, this type of analysis has been treated as a system identification problem and the analysis performed using Kalman filters. More recently, neural networks have also been used to solve the fault isolation problem.

The use of Kalman filters for fault isolation and assessment in relative engine performance diagnostics started in the late 1970s and early 1980s ([1,2]). These methods used the predictor-corrector property of the Kalman filter to predict the engine module efficiencies, flow capacities and areas, as well as sensor errors from sensor measurement deltas from a baseline. Details about the setup of the Kalman filter equations and the various matrices are discussed in several papers (Volponi and Urban [3], Doel [4], Stamatis [5], Merrington [6], Luppold [7], Gallops [8], and Kerr [9]).

More recently, neural networks have been used for gas turbine fault detection and isolation. While the Kalman filter literature focuses on long-term deterioration of the engine modules, the neural network literature focuses on single-fault isolation following a step or rate change in engine gas path measurements (DePold [10]). Typically, a step or rate change is a precursor to a reliability event causing an in-flight shut down, delay and cancellations, or unscheduled engine removals. Such trend changes in gas path measurements can be detected by a powerplant engineer through observation and interpretation, as well as by statistical and neural methods ([10]). Once a trend change has been detected, a trained neural network can isolate the engine fault ([10,11]). The Kalman filter can also be configured to isolate single faults following a trend change (Volponi et al. [12]).

In this paper, we focus on the subset of the performance diagnostics problem involving isolation of the faulty module in the gas turbine once a trend shift in one or more of the gas path measurements has been detected. Knowledge of the faulty modules reduces maintenance costs as only the faulty module needs to be opened and inspected, and not the whole engine. It is assumed that

only one module is defective. The analysis can be looked on as a way to automate the fingerprint charts to perform module fault isolation.

Fingerprint charts provided by jet engine manufacturers are routinely used by airlines for gas turbine performance diagnostics. These charts summarize the relationships between measurement deviations of gas path parameters from a baseline model to an engine fault. Experienced powerplant engineers can often look at a given trend shift in the measurement deviations and isolate the faulty module or component using the fingerprint charts. However, automation of the process allows the isolation to be performed as each measurement data point becomes available and may prevent expensive maintenance events.

Typical gas path measurements are exhaust gas temperature (EGT), low-spool rotor speed (N1), high-spool rotor speeds (N2), and fuel flow (WF). These four measurements are often called the four basic parameters and the instrumentation to measure them is available on most new and old engines. Some newer engines have additional instrumentation. However, for a fault isolation system to be widely applicable, it should be able to function with only four measurements.

Neural networks often use backpropagation learning, which is computationally intensive. Kalman filters need considerable numerical linear algebra in the form of matrix inversion, multiplication, and addition. Basically, for single-fault isolation, the Kalman filter and neural networks are being used in an attempt to automate the fingerprint charts. These fingerprint charts can be interpreted as rules, and expert systems appear to be the computationally cost effective way to automate them. In this paper, we automate the fingerprint charts using a fuzzy logic-based expert system. Fuzzy systems have been widely used in medical diagnostics (Adlassing [13] and Tazaki [14]). Recently, it has been proved that classical feedforward neural networks of the type used in engine diagnostics can be approximated to an arbitrary degree of accuracy by a fuzzy logic system, without having to go through the laborious training process needed by a neural network ([15]). In addition, the fuzzy system uses heuristic knowledge and does not need the laborious training needed by neural networks ([16]).

Module Faults

Typically, a twin-spool gas turbine engine has five modules: fan (FAN), low-pressure compressor (LPC), high-pressure compressor (HPC), high-pressure turbine (HPT), and low-pressure turbine (LPT). Most damages to the engine manifest themselves as changes in either the module efficiency or flow capacity/area. The FAN, LPC, and HPC modules have efficiencies and flow capacities associated with them, while the HPT and LPT modules have

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, New Orleans, LA, June 4–7, 2001; Paper 01-GT-013. Manuscript received by IGTI, December 2000, final revision, March 2001. Associate Editor: R. Natole.

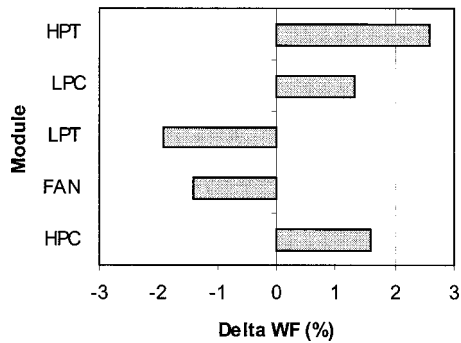


Fig. 1 Fingerprint chart showing change in fuel flow (WF) for a module efficiency decrease of -2%

efficiencies and areas associated with them. The typical fingerprint chart relating changes in measurement deltas for the four basic parameters with the faulty module is shown in Figs. 1–4. The influence coefficients used to obtain the fingerprint charts are given in Volponi et al. [12]. The fingerprint chart results assumed the following couplings between module efficiencies and flow capacities ([12]):

- 1 FAN coupled FAN (-2% η , -2.5 FC)
- 2 LPC coupled LPC (-2% η , -2.2% FC)
- 3 HPC coupled HPC (-2% η , -1.6 FC)
- 4 HPT coupled HPT (-2% η , 1.5% FP4)
- 5 LPT coupled LPT (-2% η , $+3.3\%$ FP45)

Each fault is modeled as a 2% decrease in efficiency from the baseline “good” engine. Since fingerprint charts are derived from influence coefficients, they are only approximately correct because they do not account for uncertainties in the measurement process. We will show how a fuzzy logic expert system can isolate the module faults from snapshots of measurement deltas, while accounting for uncertainty.

Fuzzy Logic System

A fuzzy logic system is a nonlinear mapping of an input feature vector into a scalar output ([17]). Fuzzy set theory and fuzzy logic provide the framework for the nonlinear mapping. Fuzzy logic systems have been widely used in engineering applications, because of the flexibility they offer designers and their ability to handle uncertainty. A fuzzy logic system can be expressed as a linear combination of fuzzy basis functions and is a universal function approximator. Further information on fuzzy logic systems is available from textbooks ([17]).

A typical multi-input single-output (MISO) fuzzy logic system performs a mapping from $V \in R^m$ to $W \in R$ using four basic components: rules, fuzzifier, inference engine, and defuzzifier. Here

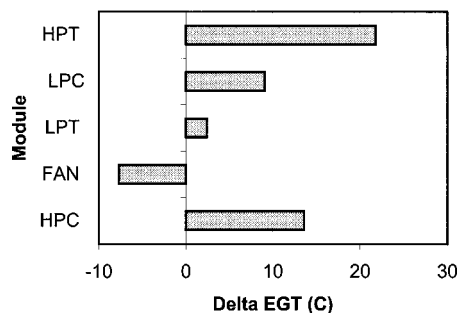


Fig. 2 Fingerprint chart showing change in exhaust gas temperature (EGT) for a module efficiency decrease of -2%

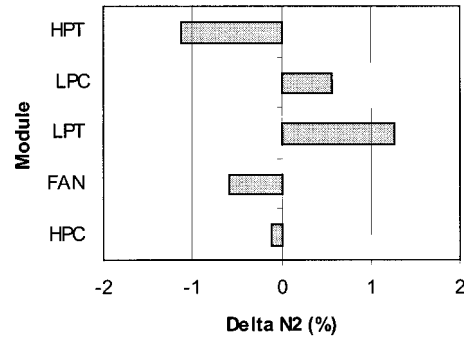


Fig. 3 Fingerprint chart showing change in high-spool rotor speed (N2) for a module efficiency change of -2%

$$f: V \in R^m \rightarrow W \in R$$

where $V = V_1 \times V_2 \times \dots \times V_n \in R^m$ is the input space and $V \in R$ is the output space.

A typical fuzzy logic system maps crisp inputs to crisp outputs using four basic components: rules, fuzzifier, inference engine, and defuzzifier. Once the rules driving the fuzzy logic system have been fixed, the fuzzy logic system can be expressed as a mapping of inputs to outputs.

Rules can come from experts or can be obtained from numerical data. In either case, engineering rules are expressed as a collection of IF-THEN statements such as “IF u_1 is HIGH, and u_2 is LOW, THEN v is LOW.” To formulate such a rule we need an understanding of

1. linguistic variables versus numerical values of a variable (e.g., HIGH versus 3.5%)
2. quantifying linguistic variables (e.g., u_1 may have a finite number of linguistic terms associated with it, ranging from NEGLIGIBLE to VERY HIGH), which is done using fuzzy membership functions
3. logical connections between linguistic variables (e.g., AND, OR etc.), and
4. Implications such as “IF A THEN B.” We also need to understand how to combine more than one rule.

The fuzzifier maps crisp input numbers into fuzzy sets. It is needed to activate rules that are expressed in terms of linguistic variables. An inference engine of the fuzzy logic system maps fuzzy sets to fuzzy sets and determines the way in which the fuzzy sets are combined. In several applications, crisp numbers are needed as an output of the fuzzy logic system. In those cases, a defuzzifier is used to calculate crisp values from fuzzy values.

Fuzzy Sets. A fuzzy set F is defined on a universe of discourse U and is characterized by a degree of membership $\mu(x)$,

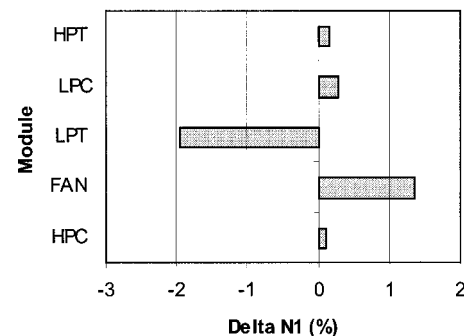


Fig. 4 Fingerprint chart showing change in low-spool rotor speed (N1) for a module efficiency change of -2%

which can take on values between 0 and 1. A fuzzy set generalized the concept of an ordinary set whose membership function only takes two values, zero and unity.

Linguistic Variables. A linguistic variable u is used to represent the numerical value x , where x is an element of U . A linguistic variable is usually decomposed into a set of terms $T(u)$, which cover its universe of discourse.

Membership Functions. The most commonly used shapes for membership functions $\mu(x)$ are triangular, trapezoidal, piecewise linear, or Gaussian. The designer selects the type of membership function used. There is no requirement that membership functions overlap. However, one of the major strengths of fuzzy logic is that membership functions can overlap. Fuzzy logic systems are robust because decisions are distributed over more than one input class. For convenience, membership functions are normalized to one so they take values between 0 and 1, and thus define the fuzzy set.

Inference Engine. Rules for the fuzzy system can be expressed as

R_i : IF x_1 is F_1 AND x_2 is F_2 AND ... AND x_m is F_m THEN $y = C_i$ $i = 1, 2, 3 \dots M$

where m and M are the number of input variables and rules, x_i and y are the input and output variables, and $F_i \in V_i$ and $C_i \in W$ are fuzzy sets characterized by membership functions $\mu_{F_i}(x)$ and $\mu_{C_i}(y)$, respectively. Each rule can be viewed as a fuzzy implication $F_1 \times \dots \times F_m \rightarrow C_i$, which is a fuzzy set in $V \times W = V_1 \times V_2 \times \dots \times V_m \times W$ with membership function given by

$$\mu_{R_i}(x, y) = \mu_{F_1}(x_1) * \mu_{F_2}(x_2) * \dots * \mu_{F_m}(x_m) * \mu_{C_i}(y)$$

where $*$ is the T -norm with $x = [x_1 x_2 \dots x_m] \in V$ and $y \in W$. This sort of rule covers many applications. The algebraic product is one of the most widely used T -norms in applications, and leads to a product inference engine. In pattern recognition problems the outputs are often crisp sets, and $\mu_{C_i}(y) = 1$ is often used for the product inference formula ([18]).

Defuzzification. Popular defuzzification methods include maximum matching and centroid defuzzification. While centroid defuzzification is widely used for fuzzy control problems where a crisp output is needed, maximum matching is often used for pattern matching problems where we need to know the output class. Suppose there are K fuzzy rules and among them, K_j rules ($j = 1, 2, \dots, L$ and L is the number of classes) produce class C_j . Let D_p^i be the measurements of how the p th pattern matched the antecedent conditions (IF part) of the i th rule, which is given by the product of membership grades of the pattern in the regions which the i th rule occupies

$$D_p^i = \prod_{l=1}^m \mu_{li}$$

where m is the number of inputs and μ_{li} is the degree of membership of measurement l in the fuzzy regions that the i th rule occupies. Let $D_p^{\max}(C_j)$ be the maximum matching degree of the rules (rules j_l , $l = 1, 2, \dots, K_j$) generating class C_j

$$D_p^{\max}(C_j) = \max_{l=1}^{K_j} D_p^{j_l}$$

then the system will output class C_{j^*} provided that

$$D_p^{\max}(C_{j^*}) = \max_j D_p^{\max}(C_j).$$

If there are two or more classes which achieve the maximum matching degree, we will select the class which has the largest number of fired fuzzy rules (a fired rule has a matching degree of greater than zero) ([19]).

Problem Formulation

Input and Output. Inputs to the fuzzy logic systems are measurement deltas and outputs are engine module faults. We have four measurements represented by $\mathbf{z}$ and five engine faults represented by $\mathbf{x}$. The objective is to find a functional mapping between $\mathbf{z}$ and $\mathbf{x}$. Mathematically this can be represented as

$$\mathbf{x} = F(\mathbf{z})$$

where $\mathbf{x} = \{\text{FAN, LPC, HPC, HPT, LPT}\}^T$ and $\mathbf{z} = \{\Delta\text{EGT}, \Delta\text{WF}, \Delta\text{N2}, \Delta\text{N1}\}^T$. Each measurement delta has uncertainty.

Fuzzification. Here FAN, LPC, HPC, HPT, and LPT are fuzzy sets denoting the five engine modules. Each fuzzy set has degrees of membership ranging from zero to one. In this paper, we are only interested in the module containing the fault. Therefore, we do not further decompose the module fuzzy sets using linguistic variables.

The measurement deltas ΔEGT , ΔN1 , ΔN2 , and ΔWF are also treated as fuzzy variables. To get a high degree of resolution, they are further split into linguistic variables. For example, consider ΔEGT as a linguistic variable. It can be decomposed into a set of terms $T(\Delta\text{EGT}) = \{\text{very high-, high-, medium-, low-, negligible, low+, medium+, high+, very high+}\}$ where each term in $T(\Delta\text{EGT})$ is characterized by a fuzzy set in the universe of discourse $U(\Delta\text{EGT}) = \{-25\text{C}, 25\text{C}\}$, which is selected to include values in the vicinity of the fingerprint chart (Fig. 2).

The other three measurement deltas are defined using the same set of terms as ΔEGT , spanning the following universes of discourse: $U(\Delta\text{WF}) = \{-4\%, 4\%\}$; $U(\Delta\text{N2}) = \{-3\%, 3\%\}$; $U(\Delta\text{N1}) = \{-3\%, 3\%\}$. Since the influence coefficients on which the fingerprints are based represent a linear model, the diagnostic system should be limited to small measurement deltas. In addition, measurement deltas larger than covered by the universe of discourse will represent a large fault indicative of a catastrophic failure.

Fuzzy sets with Gaussian membership functions are used. These fuzzy sets can be defined using the following equation:

$$\mu(x) = e^{-0.5(x-m/\sigma)^2}$$

where m is the midpoint of the fuzzy set and σ is the uncertainty (standard deviation) associated with the variable. Table 1 gives the linguistic measure associated with each fuzzy set and the midpoint of the set for each measurement delta. The midpoints are selected to span the region ranging from a perfect engine (all measurement deltas are zero) to one with significant damage.

The fuzzy set corresponding to "very high" are defined slightly differently to account for the open-ended nature of the linguistic variable.

$$\mu(x) = e^{-0.5(x-m/\sigma)^2} m_{VH-} < x \text{ OR } x < m_{VH+}$$

$$\mu(x) = 1 \quad m_{VH+} < x \text{ OR } x < m_{VH-}$$

Table 1 Gaussian fuzzy sets

Linguistic Measure (Symbol)	Midpoints		
	EGT (°C)	WF (%)	N1 and N2 (%)
Very High- (VH ⁻)	-20	-3.0	-2.0
High- (H ⁻)	-15	-2.25	-1.5
Medium- (M ⁻)	-10	-1.5	-1.0
Low- (L ⁻)	-5	-0.75	-0.5
Negligible (N)	0	0	0
Low+ (L ⁺)	5	0.75	0.5
Medium+ (M ⁺)	10	1.5	1.0
High+ (H ⁺)	15	2.25	1.5
Very High+ (VH ⁺)	20	3.0	2.0

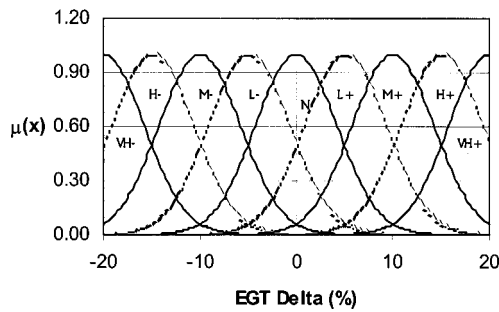


Fig. 5 Gaussian functions defining linguistic measures for fuzzy set for exhaust gas temperature (EGT) delta

Here m_{VH+} represents the midpoint corresponding to fuzzy set $VH+$. The standard deviations for ΔEGT , $\Delta N1$, $\Delta N2$, and ΔWF are 4.23C, 0.25%, 0.17%, and 0.50%, respectively, and are representative of airline data ([11]).

Figures 5–8 shows the membership functions for each of the nine fuzzy sets for ΔEGT , $\Delta N1$, $\Delta N2$, and ΔWF , respectively. There is more overlap between the fuzzy sets partitioning ΔEGT and ΔWF because these measurements have greater uncertainty, compared to $\Delta N1$ and $\Delta N2$.

Rules and Fault Isolation. Rules for the fuzzy system are obtained by fuzzification of the numerical values in the fingerprint charts using the following procedure (Wang [20] and Abe [21]):

1. A set of four measurement deltas corresponding to a given module fault is input to the FLS and the degree of membership of the elements of ΔEGT , ΔWF , $\Delta N2$, and $\Delta N1$ are obtained. Therefore, each measurement has nine degree of memberships based on the linguistic measures in Table 1.
2. Each measurement delta is then assigned to the fuzzy set with the maximum degree of membership.

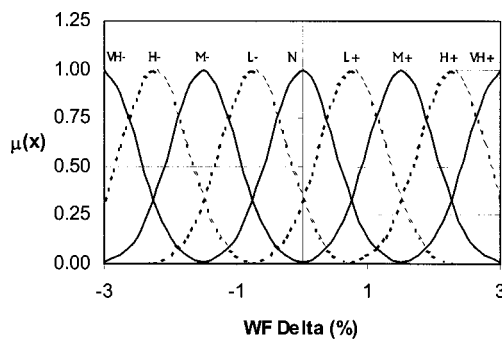


Fig. 6 Gaussian functions defining linguistic measures for fuzzy sets for fuel flow (WF) delta

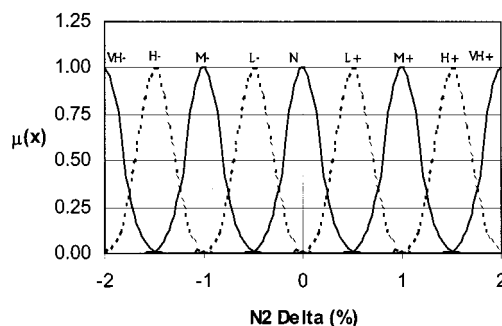


Fig. 7 Gaussian functions defining linguistic measures for fuzzy sets for high-speed rotor speeds (N2) delta

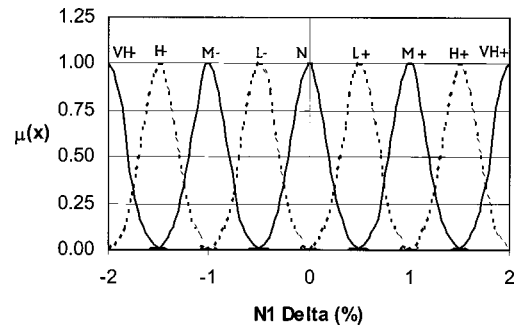


Fig. 8 Gaussian functions defining linguistic measures for fuzzy set for low-spool rotor speeds (N1) delta

3. One rule is obtained for each module fault by relating the measurement deltas with maximum degree of membership to a module fault.

These rules are tabulated in Table 2. The linguistic symbols used in this table are defined in Table 1.

These rules can be read as follows for the FAN module:

IF
 ΔEGT is medium⁻ AND
 ΔWF is medium⁻ AND
 $\Delta N2$ is low⁻ AND
 $\Delta N1$ is high⁺
 THEN
 problem in FAN module.

The rules for the other modules can be similarly interpreted. These rules provide a knowledge base and represent how a human engineer would interpret data to isolate an engine fault using fingerprint charts.

For any given input set of measurement deltas, the fuzzy rules are applied using product implication. Once the fuzzy rules are applied for a given measurement, we have degree of memberships for FAN, LPC, HPC, HPT, and LPT.

For fault isolation, we are interested in the most likely fault. The fault with the highest degree of membership is selected as the most likely fault. The fault with the second highest degree of membership is selected as the second most likely fault.

Table 2 Rules for fuzzy system

Measurement/ Module	ΔEGT	ΔWF	$\Delta N2$	$\Delta N1$
FAN	M ⁻	M ⁻	L ⁻	H ⁺
LPC	M ⁺	M ⁺	L ⁺	L ⁺
HPC	H ⁺	M ⁺	N	N
HPT	VH ⁺	H ⁺	M ⁻	N
LPT	N	H ⁻	H ⁺	VH ⁻

Table 3 Fault isolation results from fuzzy system ($\eta = -2\%$)

Module/Isolation	No. 1 choice (%)	No. 2 choice (%)
FAN	100	100
LPC	97	100
HPC	94	100
HPT	100	100
LPT	100	100
Average Success Rate	98.2	100

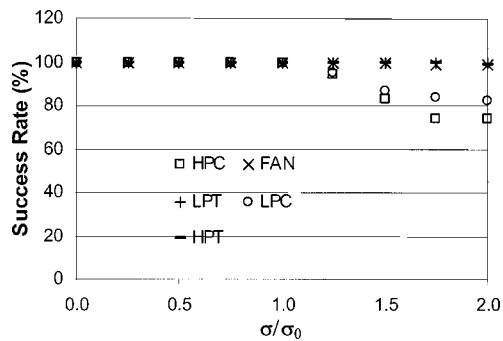


Fig. 9 Success rate in fault isolation for uncertainty levels different from design point ($= 0$)

Results and Discussion

The fuzzy system is tested using simulated data developed from the fingerprint charts. For each module, 100 data sets are generated for module faults with 2% deterioration in efficiency. Noise is added to the simulated measurement deltas using the typical standard deviations for ΔEGT , $\Delta N1$, $\Delta N2$, and ΔWF as 4.23C, 0.25%, 0.17%, and 0.50%, respectively.

Table 3 shows the results from the fuzzy system. For the FAN, LPT, and HPT modules, there is a 100% success rate with the first choice of the fuzzy system. For the LPC and HPC, there is some confounding and a slightly lower accuracy of 97% and 94%, respectively. The average success rate for the fuzzy logic system is 98.2% when only the first choice is considered.

Whenever the LPC module is not identified correctly as the first choice, the fuzzy system confuses it with the HPC module. Whenever the HPC module is not correctly identified as the first choice, the fuzzy system confuses it with the LPC module. However, if the first two choices are taken, the accuracy of the fuzzy system is 100%. The confusion between these compressor modules is because of the similarity in the directions of the fingerprints, which can be seen in the fingerprint charts as well as the fuzzy rules. In the cases where the random error is high, the fingerprints for the LPC look very similar to those of the HPC, and vice versa.

It is clear that the fuzzy logic system is able to identify the correct fault despite the presence of considerable uncertainty in the measurements. In cases where it is confounded, a human expert would also be likely to be confounded.

The fuzzy logic system was designed and tested for measurement data at a specific standard deviation, which for ΔEGT , $\Delta N1$, $\Delta N2$, and ΔWF are 4.23C, 0.25%, 0.17% and 0.50%, respectively. To observe how the fuzzy logic system performs with data at other standard deviations, we define the baseline standard deviation as $\sigma_0 = (4.23C, 0.25\%, 0.17\%, 0.50\%)$ and test the fuzzy logic system with data at standard deviation values which are multiples of

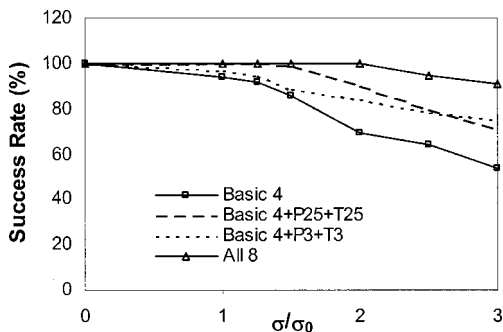


Fig. 10 Success rate in high-pressure compressor (HPC) fault isolation with additional sensors and increasing levels of uncertainty

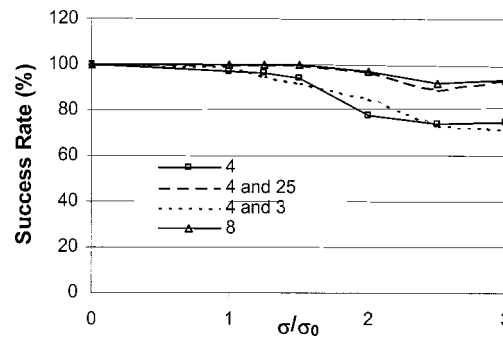


Fig. 11 Success rate in low-pressure compressor (LPC) fault isolation with additional sensors and increasing levels of uncertainty

σ_0 . The success rate for these tests (first choice only) is shown as the scatter in the data (σ/σ_0) is increased from 0 to 2 (Fig. 9). It is clear that the fuzzy logic system shows a slow deterioration in performance as the scatter in the data increases. The problems in fault isolation are mostly in the confounding between the LPC and HPC modules, and the other module faults are isolated well even with data with scatter as high as $2\sigma_0$.

Additional Measurements. Some newer engines have additional instrumentation beyond the four basic (cockpit) instruments we have considered until now. These include pressure and temperature sensors between the LPC and HPC (P25 and T25 sensors), and before the burner (P3 and T3). We next consider the effect on fault isolation as these additional sensors are included. The fuzzy logic system is extended to account for six and eight measurements. Influence coefficients provided in Volponi [12] are used for the numerical results.

Figure 10 and 11 shows the success rate in fault detection for the HPC and LPC module faults, respectively, with increasing noise levels for four, six, and eight sensors. Two cases with six sensors are considered. The first case involves the basic four and P25 and T25 sensors. The second case involves the basic four and P3 and T3 sensors. Thus there are four sensor suites: "basic 4," "basic 4 + P25 + T25," "basic 4 + P3 + T3," and "all 8." The fault isolation for FAN, HPT, and LPT are almost 100% even at high levels of uncertainty and are not shown.

For the HPC fault, the addition of the P25 and T25 sensor results in some improvement in fault isolation compared to four measurements, especially at higher levels of uncertainty (Fig. 10). Similar results are also observed with the additional P3 and T3 sensors. However, sensors at "25" location are appear somewhat more important than at "3" location for HPC fault isolation. When all four additional sensors (P25, T25, P3, and T3) are used, there is a considerable increase in fault isolation accuracy at high un-

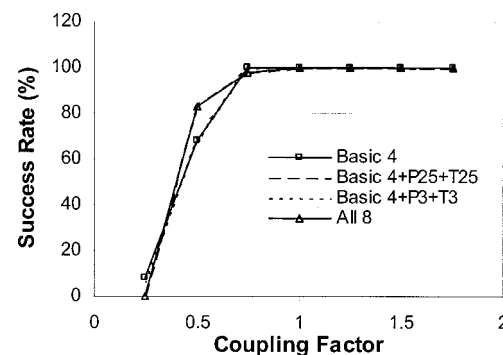


Fig. 12 Impact of coupling factor on fan (FAN) fault isolation success rate with different sensor suites

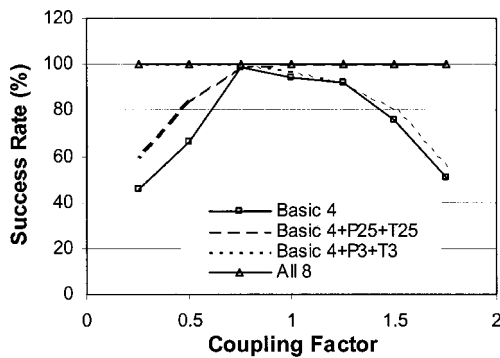


Fig. 13 Impact of coupling factor on high-pressure compressor (HPC) fault isolation success rate with different sensor suites

certainty levels. For the LPC fault (Fig. 11), the P25 and T25 sensors improve fault isolation and the P3 and T3 appear to be less important. Most of the improvement in fault isolation resulting from all eight sensors appears to come from the P25 and T25 sensors.

The fuzzy logic system is developed with implanted module faults from fingerprint charts that have preset module couplings between efficiencies and flow capacities. The FLS is tested with fault data at different coupling factors compared to the design conditions and the results for the FAN, HPC, and LPC are shown in Figs. 12, 13, and 14, respectively. For these cases, uncertainty levels are kept at the design condition ($\sigma = \sigma_0$). Again, the influence of additional sensors on fault isolation accuracy is investigated.

For the FAN module results shown in Fig. 12, using all eight sensors yields only a small influence on fault isolation robustness relative to the coupling factor. The HPT and LPT modules show negligible influence of coupling factor on fault isolation accuracy and are not shown. Figures 13 and 14 show the influence of coupling factor on the HPC and LPC module faults.

For the HPC fault, using eight measurements instead of four results in considerable improvement in robustness of fault isolation. Adding P25 and T25 instrumentation to the four basic measurements improves the fault isolation at higher coupling ratios, while adding P3 and T3 instrumentation to the four basic measurement results in improvements at lower coupling ratios.

For the LPC fault shown in Fig. 14, there is a moderate improvement in robustness of fault isolation when all eight measurements are used, relative to four measurements. Most of this improvement comes from using the P25 and T25 sensors. Thus we observe that the HPC and LPC exhibit greater robustness to modeling assumptions with the eight sensor suite. For the HPC, this robustness comes from all four extra measurements (P25, T25, P3,

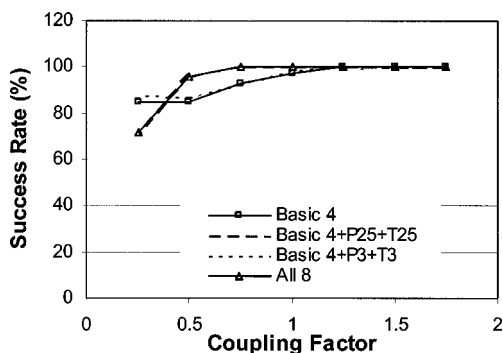


Fig. 14 Impact of coupling factor on low-pressure compressor (LPC) fault isolation success rate with different sensor suites

Table 4 Fault isolation results from fuzzy system—off-design condition ($\eta = -3\%$)

Module/Isolation	No. 1 choice (%)	No. 2 choice (%)
FAN	100	100
LPC	100	100
HPC	89	100
HPT	100	100
LPT	100	100
Average Success Rate	97.8	100

and T3) while for the LPC it comes mostly from P25 and T25 measurements.

Off-Design Condition. Finally, we consider a case at an off-design condition. The fuzzy logic system was developed based on implanted module faults with magnitude of 2% decrease in efficiency. The FLS developed with 2% implanted faults is tested with 3% implanted faults and the results are shown in Table 4.

Concluding Remarks

A fuzzy logic system is developed for gas turbine engine performance diagnostics. It takes measurement deviations from a baseline model of a “good” engine and isolates the faulty module. The fuzzy logic system operates with four basic (cockpit) measurements (exhaust gas temperature (EGT), fuel flow (WF), and low (N1) and high-spool rotor speeds (N2)) and analyzes deterioration in five modules (FAN, LPC, HPC, HPT, LPT). These measurements are available on most engines. The fuzzy logic system is based on fingerprint charts provided by engine manufacturers and used by airline engineers.

Results show that the fuzzy logic system has a success rate of almost 100% in isolating the faulty engine module with four cockpit measurements. In cases where the fuzzy logic system is confounded, it was due to large uncertainty in the data and an airline engineer may have been similarly confounded. The fuzzy logic system therefore can be used as an expert system for automating the process of interpreting gas turbine performance fingerprint charts.

Additional pressure and temperature sensors between the compressors and before the burner (P25, T25, P3, and T3) improve the fault isolation accuracy, especially as uncertainty in the measurement data increases. Additional sensors also increase the robustness of the compressor modules relative to modeling assumptions about couplings between module efficiencies and flow capacities. Therefore, additional sensors besides the four cockpit sensors are useful for accurate and robust fault isolation.

The fuzzy logic system is sufficiently robust and performs well for fault sizes that considerably different from the implanted faults used to develop the fuzzy rule base.

Nomenclature

- EGT = exhaust gas temperature
- FC = flow capacity
- FP4 = high-pressure turbine area
- FP45 = low-pressure turbine area
- N1 = low rotor speed
- N2 = high rotor speed
- T = set of terms
- U = universe of discourse of fuzzy set
- WF = fuel flow
- x = element of fuzzy set
- $\mathbf{x}$ = module faults
- $\mathbf{z}$ = measurement deltas
- Δ = change from baseline “good” engine
- η = efficiency
- $\mu_A(x)$ = degree of membership of x in fuzzy set A
- σ = uncertainty as standard deviation

References

- [1] Urban, L. A., 1972, "Gas Path Analysis Applied to Turbine Engine Conditioning Monitoring," AIAA/SAE Paper 72-1082.
- [2] Volponi, A., 1983, "Gas Path Analysis: An Approach to Engine Diagnostics," *Time-Dependent Failure Mechanisms and Assessment Methodologies*, Cambridge University Press, Cambridge, UK.
- [3] Volponi, A. J., and Urban, L. A., 1992, "Mathematical Methods of Relative Engine Performance Diagnostics," SAE Trans., **101**; Journal of Aerospace Technical Paper 922048.
- [4] Doel, D. L., 1994, "TEMPER—A Gas Path Analysis Tool for Commercial Jet Engines," ASME Paper 92-GT-315.
- [5] Stamatis, A., Mathioudakis, K., Berios, K., and Papailiou, K. D., 1991, "Jet Engine Fault Detection with Discrete Operating Points Gas Path Analysis," J. Propulsion, **7**(6), pp. 1043–1048.
- [6] Merrington, G. L., 1993, "Fault Diagnosis in Gas Turbines Using a Model Based Technique," ASME Paper 93-GT-13.
- [7] Luppold, R. H., Roman, J. R., Gallops, G. W., and Kerr, L. J., 1989, "Estimating In-Flight Engine Performance Variations Using Kalman Filter Concepts," AIAA Paper 89-2584.
- [8] Gallops, G. W., and Bushman, M. A., 1992, "In-Flight Performance Diagnostic Capability of an Adaptive Engine Model," AIAA Paper 92-3746.
- [9] Kerr, L. J., Nemec, T. S., and Gallops, G. W., 1991, "Real-Time Estimation of Gas Turbine Engine Damage Using a Control Based Kalman Filter Algorithm," ASME Paper 91-GT-216.
- [10] De Pold, H., and Gass, 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.
- [11] Lu, P. J., Hsu, T. C., Zhang, M. C., Zhang, J., 2000, "An Evaluation of Engine Fault Diagnostics Using Artificial Neural Networks," ASME J. Eng. Gas Turbines Power, **123**, pp. 240–246.
- [12] Volponi, A. J., DePold, H., Ganguli, R., and Daguang, C., 2000, "The Use of Kalman Filter and Neural Network Methodologies in Gas Turbine Performance Diagnostics: A Comparative Study," ASME Paper 00-GT-547.
- [13] Adlassing, K. D., 1986, "Fuzzy Set Theory in Medical Diagnostics," IEEE Trans. Syst. Man Cybern., **16**, pp. 260–264.
- [14] Tazaki, E., et al., 1986, "Development of Automated Health Testing and Services System via Fuzzy Reasoning," *Proc. IEEE Int. Conf. On Systems, Man and Cybernetics*.
- [15] Hong, X. L., and Chen, P. C. L., 2000, "The Equivalence between Fuzzy Logic Systems and Feedforward Neural Networks," IEEE Trans. Neural Netw., **11**(2), pp. 103–111.
- [16] Zadeh, L., 1996, "Fuzzy Logic=Computing With Words," IEEE Trans. Fuzzy Syst., **4**(2), pp. 103–111.
- [17] Kosko, B., 1997, *Fuzzy Engineering*, Prentice-Hall, Englewood Cliffs, NJ.
- [18] Chi, Z., Yan, H. M., and Pham, T., 1998, *Fuzzy Algorithms: With Applications to Image Processing and Pattern Recognition*, World Scientific, Singapore.
- [19] Chi, Z., and Yan, H., 1993, "ID3 Derived Fuzzy Rules and Optimized Defuzzification for Handwritten Character Recognition," IEEE Trans. Fuzzy Syst., **1**(1), pp. 24–31.
- [20] Wang, L. X., and Mendel, J. M., 1992, "Generating Fuzzy Rules by Learning From Examples," IEEE Trans. Syst. Man Cybern., **22**(6), pp. 1414–1427.
- [21] Abe, S., and Lan, M. S., 1995, "A Method for Fuzzy Rules Extraction Directly From Numerical Data and Its Application to Pattern Recognition," IEEE Trans. Fuzzy Syst., **3**, pp. 18–28.

Interpretation of Weighted-Least-Squares Gas Path Analysis Results

D. L. Doel

GE Aircraft Engines,
One Neumann Way,
Evendale, OH 45215-6301
e-mail: dave.doel@ae.ge.com

Gas turbine performance analysis programs convert raw measurements (thrust, fuel flow, gas path temperatures, and pressures) into engine and component figures of merit (e.g., specific fuel consumption, component efficiencies). The derived parameters allow a diagnostician to recognize a poorly performing engine and help to identify the root cause for the problem. These goals can be foiled by measurement problems or by analysis assumptions embedded in the programs. A gifted analyst learns to recognize measurement or assumption problems via characteristic "fingerprints" in the program output. The latest data analysis tools tackle the same issues using a weighted-least-squares algorithm which estimates sensor errors at the same time component health is being determined. The weighted-least-squares algorithm modifies the "fingerprint" of a measurement problem and also that of a problem-free solution. To develop intuition for these new programs, the analyst must understand how weighted-least-squares works and how normal and problem situations are reflected in the output. This paper describes the weighted-least-squares algorithm, as implemented in GE's TEMPER program, with the goal of stimulating a new intuition for interpreting gas path analysis results. [DOI: 10.1115/1.1582492]

Introduction

Data analysis seeks to determine the health of an engine and, when a performance deficiency is found, to identify the cause(s) for that deficiency. Overall performance measures such as specific fuel consumption and exhaust gas temperature (EGT) are used to monitor engine health. Individual components are evaluated by computing their efficiency and their flow capacity (also called pumping). For most engines, the compression efficiencies can be computed directly from pressure and temperature measurements. Efficiencies of the turbines must be deduced indirectly due to the lack of inlet or outlet temperature or pressure data.

The indirect calculations are based upon assumptions about the engine. Typically, the high pressure turbine flow function is assumed known (and fixed), allowing calculation of the turbine inlet temperature from fuel flow. High pressure turbine inlet pressure is computed from compressor discharge pressure assuming a known pressure loss through the combustor. The flow through the high pressure turbine is established as part of the turbine inlet temperature calculation. This, together with some assumptions about parasitic flows, permits calculation of the low pressure turbine inlet temperature (the calculation equates high pressure turbine work to high pressure compressor work). If low pressure turbine inlet pressure is available, low pressure turbine flow function can be calculated. Otherwise, LPT flow function is assumed and used to calculate the inlet pressure. Low pressure turbine exit pressure and temperature are computed using nozzle continuity, and equality of fan/low pressure compressor work to low pressure turbine work, respectively.

Surprisingly, this approach creates one (or two) redundant measurements, low pressure turbine inlet temperature, also called EGT, (and LPT exit temperature when it is available). EGT is not used in the component performance calculation despite its importance to the overhaul process (EGT margin is the accepted measure of engine quality for overhauled engines).

The analysis approach just described is commonly used to analyze development test data. Development engines often have several multi-element pressure and temperature rakes at important measurement stations. The radial and circumferential profiles of these measurements are used to identify faulty probes yielding consistent, accurate plane averages at the key measurement stations. The same analysis approach was initially used with overhaul data, but problems were encountered.

Engines tested after overhaul have limited instrumentation. When temperatures and pressures are available, they usually are single element probes. Static pressure taps are more common than total pressure probes and the sensors may not be at the appropriate axial position. Often, the pressure and temperature sensors are part of the engine hardware, meaning that they have been in the engine through years of revenue service. The quality of these measurements is significantly poorer than the norm for development testing. Completely failed measurements are not uncommon. In this environment, the conventional data analysis approach may either fail (if a critical measurement is unavailable) or yield bizarre results (efficiencies above 100% for example) when measurements have shifted significantly.

Gifted analysts learn to recognize the symptoms of these problems in the analysis results. For example, if the high pressure compressor efficiency is inordinately high at the same time that high pressure turbine efficiency is low, it is reasonable to suspect a problem with the pressure or temperature measurement at the compressor exit. The temperature is the more probable fault, but this can be confirmed by examination of overall compressor pressure ratio or turbine pressure ratio (if high pressure turbine exit pressure is available). Once the measurement fault has been isolated, the analysis can be "corrected" by adjusting the errant measurement to yield more reasonable results. However, it is no longer possible to uniquely determine the compressor and turbine efficiencies because there is a continuum of valid solutions.

When more than one hard sensor fault is present, the problem becomes more difficult; even experienced analysts may not correctly identify the faulted parameters. Sensor noise adds to the difficulty. Experienced performance engineers working for the gas turbine manufacturers are often able to obtain useful results de-

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Amsterdam, The Netherlands, June 3–6, 2002; Paper No. 2002-GT-30025. Manuscript received by IGTI, December 2001, final revision, March 2002. Associate Editor: E. Benvenuti.

spite these difficulties. But the analysis programs are provided to airlines and overhaul shops whose personnel lack the experience to cope with these difficulties.

Urban [1] pioneered the use of weighted-least-squares in a program named GPA¹ as an approach to solving these problems. The algorithm can simultaneously estimate engine hardware condition and sensor errors. It achieves this by augmenting the analysis with data specifying the expected variation in the hardware parameters (efficiency and flow) and of the sensor errors. Much of the following material seeks to convey a heuristic understanding of how the algorithm achieves this task; hence, no brief explanation will be offered here. GE developed a similar tool in the late 1970s called TEMPER² (described in Doel [2]). TEMPER was developed for overhaul acceptance test data and was later extended to analysis of cruise snapshot data. The discussion here is limited to the test cell version.

The initial implementations of GPA and TEMPER were pure applications of weighted-least-squares concepts. Later, both development teams acknowledged the need to be able to identify a single large-magnitude fault (which could be a component problem or a large measurement error). These modifications to the basic algorithm are described in Doel [2] for TEMPER and Volponi [3] for GPA. The weighted-least-squares algorithm is linear—that is, the inferred component changes and sensor errors are a linear function of the sensor deviations from the reference baseline. When invoked, the large-fault logic introduces a nonlinear element (in TEMPER, a step-change between adjacent linear regions).

Both GPA and TEMPER have reported many successes. The algorithms are most frequently used on overhauled engines that fail an outbound performance test. Some users process all outbound data through one of the programs (this practice is recommended). The tools may also be used on data from inbound runs, in an effort to determine the most cost-effective work scope (this use is not really compatible with the design of the tools).

There are also instances where TEMPER and GPA are not effective. It is clear that sensor data can be so poor that no algorithm can determine the state of an engine from available measurements. There can be flaws in program setup that introduce significant problems (one such problem, incorrect baselines, is discussed later in this paper).

The weighted-least-squares algorithm behaves differently when presented pure fault data than does a conventional algorithm. The conventional algorithm will accurately report the hardware problem used to generate contrived sensor data (if the sensor data represents a 1% loss of turbine efficiency and nothing else, the conventional algorithm will report a 1% loss of turbine efficiency as the sole deviation from normal performance). TEMPER and GPA, presented with the identical input, will report a smaller loss of turbine efficiency (perhaps 0.6%) and will report other problems—a small loss of compressor efficiency, small measurement errors in EGT and fuel flow and other small component and measurement deviations. The same type of behavior occurs if a data representing a pure measurement problem is submitted to the weighted-least-squares algorithm although the spurious errors are in component efficiencies and other measurements (of course, this type of input is deadly to a conventional algorithm). In short, weighted-least-squares programs have the somewhat disconcerting feature that the user must learn how to interpret the results when the input data is without error.

Of course, real input data is not perfect, but the peculiarities of the weighted-least-squares algorithm call for changes to the way the analyst interprets the results relative to the conventional algorithms. Any significant performance problem in the engine will cause an increase in EGT and in fuel flow. GPA or TEMPER will partly explain these increases as hardware problems, but there will be a residual error attributed to the EGT and fuel flow

measurements.³ The analyst needs to be aware of this behavior when deciding on maintenance to restore engine performance. The logic to identify large faults can help this interpretation task by singling out a single large component problem (or a single large measurement error), but only if the program baselines are substantially free of flaws.

Some (Possibly Controversial) Observations Regarding Data Analysis

Socrates said, “The unexamined life is not worth living,” encouraging his pupils to step back from their daily activities to examine the deeper questions of life. It is equally useful for the data analyst to take time to ponder the strategy of his art as a means to understand the strengths and limitations thereof. This paragraph offers some observations that have suggested themselves to the author over a career devoted largely to data analysis. Needless to say, not all of these ideas are original with the author; most were offered by or derived from the observations of colleagues.

1. The purpose of data analysis is to support the making of decisions. There is little value to knowing efficiency for its own purpose. The information is used in design to determine whether a developing product can be improved. It is used in maintenance to decide whether a part continues to be serviceable. The analysis results are useful to the extent that they contribute to making a sound decision.
2. There is no information in analysis results that is not already contained in the data used to generate those results. Analysis serves two purposes: to focus attention on some specific attribute of the test item that is not readily seen in the raw data, and to allow interpretation of the measurements using a set of assumptions that the analyst (or the designer of the analysis tool) believes capable of illuminating the real behavior of the test item.
3. Assumptions filter and modify the information contained in the raw measurements. No set of assumptions is always effective. It is critical to understand the assumptions underlying the data analysis and their impact on the results. Without this knowledge, the analyst is ill equipped to recognize an erroneous solution.
4. There is no such thing as “cheating” in data analysis. Analysts often feel constrained to accept the results of a tool being used to interpret data from a device being monitored. They should reject this constraint, and use to their advantage every morsel of information which might shed light on the problem confronting them.

TEMPER Output

Thus far, the discussion has been about general weighted-least-squares algorithms. From here, the focus will be on the TEMPER algorithm described in Doel [2]. A sample TEMPER output file is shown in Fig. 1.

The output is presented on two pages. The first page shows the EGT deviation attributed to each of the major modules and to the EGT measurement error (EGT margin is the “coin of the realm” for overhaul shops). A positive value indicates that the component’s performance is better than the baseline engine; a large negative value suggests that the component is a candidate for rework. Beneath the visual display of EGT impact is a table providing the SFC impact for each of the components. The bottom of the first page reports large component deviations or measurement errors.

The second page provides details of the solution. The first table on the page presents the solution for the components. The column

¹For Gas Path Analysis

²For Turbine Engine Module Performance Estimation Routine

³It is possible to adjust the statistical inputs to a weighted-least-squares algorithm to prevent this problem, but doing so converts the analysis into a linear approximation of the conventional algorithm.

designated “PCT XHAT” gives the actual deviation (for example, Fig. 1 indicates that the low pressure turbine efficiency is 1.67% poorer than the target). The column designated “XH/SIG” expresses the same deviation in standard deviations. This indicates how unusual an observed deviation is relative to the statistical model.

The “MEASUREMENT STATUS” table in Fig. 1 reproduces the values of measurements input to the program, provides the deviations of the independent measurements from the baseline model, and lists the assessment of the sensor errors in both percent and number of standard deviations. The values of other input parameters are listed under the heading, “INPUT VARIABLES.” Finally, “PERCENT PROBABILITY” and “RESIDUAL ERROR” indicate whether the data appears to be consistent with the statistical model (more on this later).

The TEMPER Algorithm

TEMPER uses weighted-least-squares to apportion observed sensor deviations between hardware deviations (component efficiencies and flow capacities) and measurement errors. The weighted-least-squares algorithm bases the apportionment on expected variation of the hardware parameters and on the expected measurement error. In the simplest case, the variations can be expressed as standard deviations—for example, a standard deviation can be used to express the statistical variation of turbine efficiency for overhauled engines. Similarly, another standard deviation describes the statistical behavior of the fuel flow error. These standard deviations imply the expected range for turbine efficiency deviation and for fuel flow measurement error.

If the standard deviation for turbine efficiency is large and the standard deviation for fuel flow is small, TEMPER will favor a solution that interprets an indicated fuel flow deviation as a change in turbine efficiency. Conversely, if the standard deviation for fuel flow is large and the standard deviation for turbine efficiency is small, the same indicated fuel flow deviation will be interpreted as a fuel flow measurement error. Weighted-least-squares accomplishes this by minimizing an error term that is the sum of squares of individual parameter deviations divided by their respective standard deviations. If the only variables in the analysis were turbine efficiency and fuel flow error, the error term to be minimized would be

$$J = \frac{1}{2} \left[\left(\frac{\Delta \eta_T}{\sigma_{\eta_T}} \right)^2 + \left(\frac{\text{err}_{WF}}{\sigma_{\text{err}_{WF}}} \right)^2 \right]. \quad (1)$$

This provides one relationship between the two unknowns ($\Delta \eta_T$ and err_{WF}). A second equation is derived from the fuel flow deviation observed in the test, ΔWF .

$$\Delta WF = \text{err}_{WF} + \frac{\partial WF}{\partial \eta_T} \Delta \eta_T \quad (2)$$

The solution to this greatly simplified case is obtained by solving Eq. (2) for err_{WF} in terms of ΔWF and $\Delta \eta_T$, substituting the result into Eq. (1), and solving for the value of $\Delta \eta_T$ that minimizes J . The resulting expression for J is

$$J = \frac{1}{2} \left[\left(\frac{\Delta \eta_T}{\sigma_{\eta_T}} \right)^2 + \left(\frac{\left(\Delta WF - \frac{\partial WF}{\partial \eta_T} \Delta \eta_T \right)}{\sigma_{\text{err}_{WF}}} \right)^2 \right]. \quad (3)$$

The minimum J is found by solving

$$\frac{\partial J}{\partial \Delta \eta_T} = 0 \quad (4)$$

which gives

$$\Delta \eta_T = \frac{\frac{\partial WF}{\partial \eta_T} \Delta WF}{\left[\left(\frac{\sigma_{\text{err}_{WF}}}{\sigma_{\eta_T}} \right)^2 + \left(\frac{\partial WF}{\partial \eta_T} \right)^2 \right]}. \quad (5)$$

Note that this expression gives the expected limiting behavior. If $\sigma_{\text{err}_{WF}}$ is much larger than σ_{η_T} , $\Delta \eta_T$ approaches zero and $\text{err}_{WF} \approx \Delta WF$. Conversely, if $\sigma_{\text{err}_{WF}}$ is much smaller than σ_{η_T} , the solution approaches

$$\Delta \eta_T \approx \frac{\Delta WF}{\left(\frac{\partial WF}{\partial \eta_T} \right)}$$

which attributes the ΔWF to turbine efficiency ($\Delta \eta_T$).

Since the solution minimizes J , the size of that minimum provides an indication of the success of the analysis. If J is large, there is an indication that the engine is not typical of those used to generate the baseline—either it is very badly deteriorated in one or more components, there is a large measurement error, or the engine is experiencing a fault that is not addressed by the analysis. If J is small, the inference is that the engine was either close to the baseline or that the difference was easily explained using the available hardware parameters and measurement errors.

The residual error reported by TEMPER is twice the minimum value of J established via the generalized version of Eq. (4). For the simplified case, using the turbine efficiency and fuel flow errors described above, the residual error, RE, is given by the expression

$$\text{RE} = 2J = \frac{\Delta WF^2}{\sigma_{\text{err}_{WF}}^2 + \left(\frac{\partial WF}{\partial \eta_T} \sigma_{\eta_T} \right)^2}. \quad (6)$$

Equation (6) shows that the residual error increases in proportion to the square of the raw measurement deviation from the baseline (ΔWF in this case). It is reduced by the ability to explain this deviation either as a measurement error (represented by $\sigma_{\text{err}_{WF}}$) or as a deviation in a component [represented by the term, $(\partial WF / \partial \eta_T \sigma_{\eta_T})$]. A small value for residual error indicates either that the engine tested close to the expected baseline, or that its deviation from the baseline occurred in an expected way (as represented by the hardware and measurement standard deviations).

The example uses a single measurement and a single hardware fault. TEMPER is typically interpreting simultaneous deviations of approximately ten independent measurements and attempting to estimate ten hardware parameters (efficiency and flow capacity of the fan, LP compressor, HP compressor, HP turbine, and LP turbine). Ignoring correlation between faults or measurement errors, the impact on the weighted-least-squares error parameter is to add terms similar to those already shown in Eq. (1).

$$J = \frac{1}{2} \left[\left(\frac{\Delta \eta_{\text{HPT}}}{\sigma_{\eta_{\text{HPT}}}} \right)^2 + \left(\frac{\Delta \eta_F}{\sigma_{\eta_F}} \right)^2 + \left(\frac{\Delta \Gamma_F}{\sigma_{\Gamma_F}} \right)^2 + \dots + \left(\frac{\text{err}_{WF}}{\sigma_{\text{err}_{WF}}} \right)^2 + \dots \right] \quad (7)$$

Equation (7) gives a more general expression for the weighted-least-squares error parameter, ignoring correlation effects. The completely general expression for J is given by Eq. (8), which uses matrix nomenclature.

$$J = \frac{1}{2} \{ \bar{x}^T M^{-1} \bar{x} + (\bar{z} - H \bar{x})^T R^{-1} (\bar{z} - H \bar{x}) \} \quad (8)$$

The computation in TEMPER is performed according to Eq. (8). Only a few correlation terms are included in the model (only between flow and efficiency faults for the compression compo-

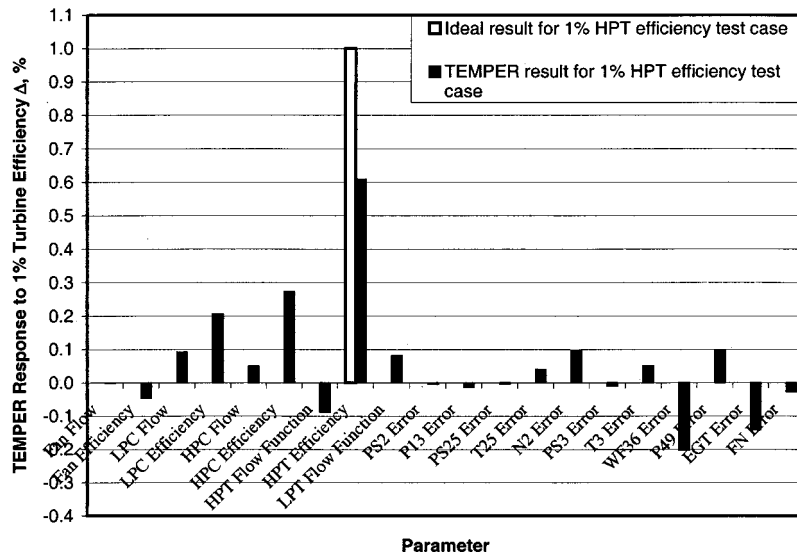


Fig. 2 Response to pure 1% HPT efficiency fault

nents and none for the measurement system). Thus, the TEMPER J differs from that of Eq. (7) only in the addition of terms to account for the correlation between fan flow and efficiency, booster flow and efficiency and HP compressor flow and efficiency. The solution to Eq. (8) (that minimizes J) is

$$\hat{x} = (M^{-1} + H^T R^{-1} H)^{-1} H^T R^{-1} \bar{z}. \quad (9)$$

Residual Error and Fault Logic

The residual error is computed by substituting $\hat{x}$ for $\bar{x}$ in Eq. (8) and multiplying the resultant J_0 by 2. Since $\hat{x}$ is the best solution, the magnitude of J (or of $2J$) indicates how well TEMPER was able to explain the observed measurement pattern. A small value for residual error means that either the engine tested close to the expected baseline or that the deviations were easily “explained” based upon the statistical model. A large residual error means that the engine tested far from the expected baseline and that the deviation did not fit any of the preferred patterns in the statistical model.

One potential source of a large residual error is a large performance deviation in a single engine component. Another is an unusually large error in one of the measurements. TEMPER explores these two possibilities through its “fault logic.” The fault logic attempts to significantly reduce the residual error by allowing a single component’s performance to be far from its expected baseline or by allowing a single measurement to have an abnormally large measurement error.

To explore the possibility that a specific measurement has a large error, the analysis is repeated with that measurement error’s standard deviation increased by a factor of 100 while all other statistical model parameters are held constant. If the measurement has a large error, the residual error from the resulting analysis will be significantly reduced from the base analysis. Conversely, if the specific measurement’s error is small, the change in the residual error will also be small.

A similar strategy is used to explore the possibility of a single component being far from its expected baseline. The standard deviation for that component’s flow and efficiency are each increased by a factor of 100 (flow and efficiency are also tested individually) with the other statistical model parameters held constant. This permits a solution where the component deviates far from its normal performance. If the component is badly deteriorated, the residual error should be significantly reduced from the base case; otherwise the change should be small.

TEMPER’s fault logic tries each measurement error and each selected component fault, one at a time, looking for the best interpretation for the observed measurement set (this is only done when the residual error is sufficiently large in the base solution). TEMPER explores some faults not included in the base case, such as parasitic flows and “setting parameter” measurement errors (e.g., fan speed or engine inlet temperature). The solution providing the greatest reduction to the residual error (actually the one that increases the probability the most) is used in place of the base solution if the fault solution probability is sufficiently large.

Interpreting a TEMPER Solution

A standard technique used to check out a conventional data reduction program is to simulate input data using a performance model (for example the measurements associated with a 1% increase of turbine efficiency), process this simulated data through the data reduction program, and confirm that the output matches the simulated performance change within expected accuracy (the program should report the expected change in turbine efficiency). The conventional program should be able to reproduce the change implied by the input data accurately so long as it computes the associated parameter. It is also useful to explore the response of the conventional program to specific measurement errors and to errors in the underlying assumptions. This information is used by an analyst to understand questionable results. For example, an unusually high turbine efficiency coupled with an unusually low compressor efficiency might be caused by an error in measured compressor discharge temperature or pressure; the response data for pure sensor errors can be used to identify the error required to produce the observed result. It can also indicate other output shifts that should have been caused by the error allowing the analyst to substantiate an interpretation of the data.

If one performs a comparable test on TEMPER (or any weighted-least-squares algorithm), the results do not conform to expected behavior. The results of such a test are shown in Fig. 2. The “ideal” solution would show a 1% increase in HP turbine efficiency and no change to any other output parameter. TEMPER reports an increase of HP turbine efficiency of slightly over 0.6% and reports changes to other parameters (most notably, a 0.2% increase in LP compressor efficiency, almost a 0.3% increase of HP compressor efficiency, and measurement errors in fuel flow and EGT). These results are inevitable for the weighted-least-squares algorithm. The pattern shown in Fig. 2 depends upon the

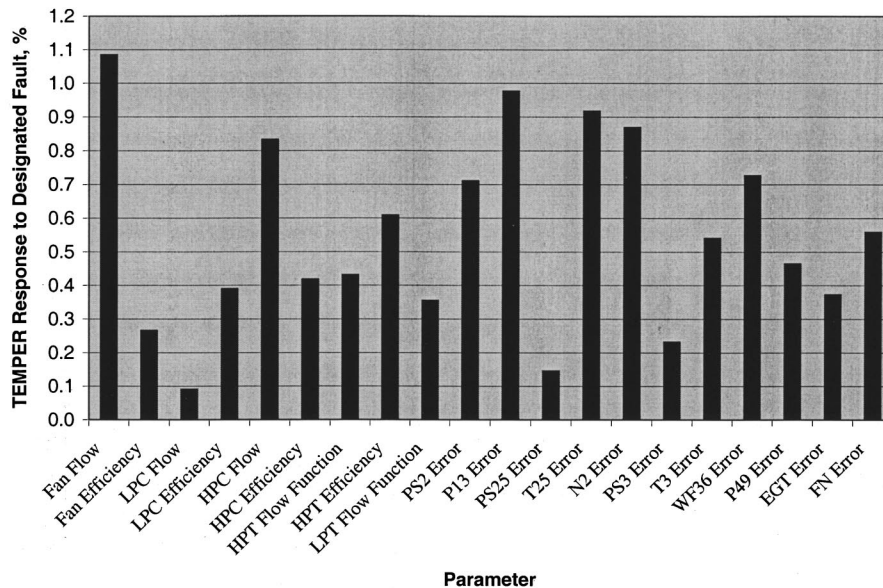


Fig. 3 Response to faults & measurement errors

statistical assumptions for the component variations, M , and for the measurement errors, R . Not every weighted-least-squares program (or even every TEMPER program) will display the pattern of Fig. 2; a change to the statistical inputs will produce a change in the output.

The weighted-least-squares algorithm is described as being maximum likelihood or minimum variance. This indicates the goal of the analysis is not to find the one, true solution, but instead to find a solution with the highest probability of being near to the one, true solution. It seeks to do this using many fewer equations (one per measurement) than it has values to find (it estimates the error for each measurement and the deviation for each hardware parameter). It accomplishes this seemingly impossible task by minimizing a sum-of-squares error term like that shown in Eq. (7). For a pure fault, such as turbine efficiency, the sum-of-squares error term is always smaller when the deviation is distributed among several outputs than when it is assigned to a single cause. This feature of the algorithm is the source of the underestimation of a single fault. The TEMPER fault logic is designed to correct this for a single fault that is large enough, but most solutions will not use the fault logic; even those that do will exhibit the behavior for parameters other than the one singled out by the fault logic. Understanding this aspect of weighted-least-squares is key to developing an intuition for interpreting the results.

When the fault logic is not invoked, TEMPER is a linear calculation. Thus a pattern, like that in Fig. 2 can be generated for every hardware variable and for every measurement error. Each pattern scales in proportion to the input until the fault logic is activated. It is also useful to explore the response characteristic for items not included in the analysis. For example, data to simulate an increase in parasitic flow can be input to observe the response to a fault not in the normal repertoire.

No analyst will assimilate the complete pattern for the 9 component variables, 11 measurement errors, and all additional parameters that might be tested, nor would such knowledge contribute to understanding a particular solution. However, it is useful to understand how much TEMPER underestimates any particular fault (for example, from Fig. 2 we learn it reports roughly 60% of the input magnitude of a pure HP turbine efficiency fault). It is also useful to understand the response to extraneous problems not part of the normal solution.

Figure 3 shows the response to the 9 component variables and 11 measurement errors. The values in Fig. 3 represent the part of

the solution that correctly responds to a pure input problem (denoted "diagonal" response). For example, the HPT efficiency response shown in Fig. 3 is the same as from Fig. 2. A display, like that in Fig. 3 can be generated for any weighted-least-squares program. The responses of the individual variables will depend on available measurements, on the component variables, and on the statistical assumptions made for component changes and measurement errors.

A calculation like that in Fig. 3 could be used to establish the statistical inputs. The designer could opt to make all component faults and measurement errors equally detectable (which would make all vertical bars in Fig. 3 have the same height). This strategy has not been used to establish the TEMPER statistical inputs, although the response characteristics are considered in establishing the values. Doel [4] describes derivation of the statistical inputs for TEMPER; the discussion will not be repeated here.

Several of the responses in Fig. 3 are surprising and deserve comment.

1. The fan flow response is greater than 100% although earlier comments suggested this could not happen. TEMPER includes a strong correlation between fan flow and fan efficiency changes which causes a pure fan flow shift to appear "unnatural." Newer response analyses address this issue by determining the response to a correlated change of flow and efficiency. This gives the expected result. The same phenomenon causes the very low fan efficiency response in Fig. 3.

2. The LPC flow response is very low (less than 10%). The available measurement set provides a relatively weak indication of LPC performance. The statistical parameters are sometimes adjusted to emphasize the components (HPC and HPT) that are known to be prime contributors to engine performance. These explain the poor response. Note that the PS25 error response is also poor.

3. The P13 error response is surprisingly strong. P13 is one of three test cell measurements which provide an indication of fan flow; the other two are PS2 (used to compute airflow) and thrust. P13 is the "weakest" of the three in the test cell but is the only one available on wing. In the test cell version of TEMPER, P13 plays a significant role only if the other measurements are unavailable or are clearly wrong.

4. The EGT measurement error response is smaller than it could be. Since EGT is the critical measure of test cell perfor-

mance, most users are not willing to accept a report that high EGT is due solely to measurement error. This is especially true since long experience shows that replacing the EGT measurement system typically has no impact on measured engine performance. The choice of statistical inputs tends to drive an EGT problem into the core engine components (HPC, HPT) and exaggerates the fuel flow measurement error. This knowledge is useful for interpreting results. A report of an EGT measurement problem may indicate combustor distress or one or more plugged fuel nozzles.

Weighted-least-squares has an altered reliance on assumptions compared to a conventional algorithm. Conventional algorithms typically select a single "path" for determining flows in the engine. A frequent choice for core flow is one of the turbine flow functions. The turbine flow functions change slowly with power setting and are also less likely to be changed by engine deterioration. Fan flow is determined using inlet total and static pressure measurements when these are available. Other possible choices for determining core flow include core nozzle continuity, and LP or HP compressor maps. Fan flow can be determined from fan nozzle continuity or from the fan map (it could also be inferred from measured thrust). The conventional programs choose a favored path (for example, LP turbine flow function continuity) and use the available data to compute HP turbine flow function, core nozzle flow coefficient, HP compressor flow capacity, etc. A similar strategy is used for fan flow.

When sufficient measurements are available, conventional programs also choose from multiple options for determination of fan/LP compressor work per pound of air and HP compressor work per pound of air. Fuel flow is generally used as a key measurement in this process which allows the programs to estimate measurement errors in the downstream temperatures (EGT, LP turbine discharge temperature). Experienced analysts use these alternate calculations (nozzle flow coefficients, temperature measurement errors, etc.) to expose measurement or assumption problems in the conventional solution. Isolation of the identified problems is done using the influence coefficients developed in the program checkout.

TEMPER does not select a path to determine core flow, fan flow, etc. Instead, it employs each of the available analysis paths, using the statistical inputs for components and measurement errors to weight each path.⁴ This use of multiple paths should improve the assessment of core flow and other cycle attributes to which it applies (at least on a relative basis).

The determination of efficiencies relies mainly on upstream and downstream measurements, with consistency checking based on overall performance measures such as fuel flow, rotor speed and EGT. For example, the overall performance will establish the approximate core engine efficiency. The relative levels of HP compressor and HP turbine efficiency will be influenced strongly by the measured compressor discharge temperature, but the impact of the temperature will be moderated by the implied changes to the two efficiencies. If the temperature tries to simultaneously raise compressor efficiency and lower turbine efficiency, TEMPER is apt to increase the measurement error and diminish the component changes.

One other difference from a conventional algorithm is that TEMPER automatically alters its analysis strategy when measurements are missing or obviously wrong. If a key measurement is missing in a conventional program, it is likely to refuse to process the data forcing some accommodation by the analyst. TEMPER may report the absence of the measurement, but is likely to present a solution as if nothing were changed. The analyst needs to recognize this so that the altered solution can be properly evalu-

ated (a good strategy is to compare the behavior of other solutions with the measurement available versus unavailable).

Probability

Earlier discussion suggested that a "small" value of residual error is generally indicative of a "good" solution and that a "large" residual error suggests an inability to interpret the observed measurement pattern using the built-in statistical model. However, the earlier discussion does not provide a quantitative meaning for "small" or "large" residual error. Bryson [5] shows that the expected value of the residual error ($2J_0$) is equal to the number of independent measurements (for a typical test cell analysis, from 8 to 12). This gives some indication of what "small" and "large" mean, but it does not provide a complete answer.

One approach to answering the question would be to perform analyses on many readings from many engines, recording the residual error from each analysis. Over time, one might observe that fewer than 5% of the cases had a residual error as large as 20 and that ~60% of the solutions had a residual error less than 10. This could suggest defining a "small" residual error as being less than 10 and that a "large" residual error as being greater than 20. This approach is equivalent to the method used to evaluate the residual error.

Specifically, TEMPER defines a parameter called "PERCENT PROBABILITY," (hereafter just probability) that is computed from the residual error, to express the fraction of the solutions that can be expected to have the same or smaller residual error. If a solution indicates that the residual error is 9.5 and the probability is 57.6, the precise meaning is that 57.6% of solutions have a residual error at least as large as 9.5.

The TEMPER probability calculation is derived from a statistical interpretation of the residual error. TEMPER assumes that the residual error ($2J_0$) follows the chi-squared distribution. This interpretation is derived from the observation by Zwillinger [6]: If $Y_1, Y_2, \dots, Y_n$ are independent and identically distributed normal random variables with a mean of 0 and a variance of 1, then $\chi^2 = \sum_{i=1}^n Y_i^2$ is distributed as chi-square with n degrees-of-freedom. Note the similarity between this expression and the expression for J given by Eq. (7). Each of the terms in Eq. (7) should have a mean of zero (because the baseline is meant to represent an "average" engine) and a standard deviation of one (because the numerator is divided by the standard deviation associated with the specific term). Equation (7) suggests that the number of degrees-of-freedom would be equal to the sum of the number of hardware parameters being estimated and the number of independent measurements, but this is for $2J$ rather than $2J_0$. The number of degrees-of-freedom is reduced by the solution for the hardware parameters to be equal to the number of measurements. Zwillinger [6] indicates that the mean of a chi-square variable is equal to the number of degrees-of-freedom. This coincides with the analysis of Bryson [5].

This discussion is not intended to be a formal mathematical proof for applying the chi-square distribution to the residual error; rather it shows the rationale for choosing the chi-square distribution to compute the probability. A formal mathematical analysis might show this result to be incorrect, but using the chi-square distribution is approximately correct, and allows for a meaningful interpretation of the residual error in terms that are more commonly understood.

The chi-squared distribution determines probability from two inputs: the residual error and the number of degrees-of-freedom. The residual error is a sum of squared errors such as computed in Eq. (8) (and doubled to be compatible with the chi-squared definition). The number of degrees-of-freedom describes the number of independent measurements included in the residual error term. It is reasonable to expect that the size of the residual error will

⁴Some items, for example core nozzle flow coefficient, have no assigned uncertainty so that their weighting is influenced by the standard deviations of measurement errors associated with the determination of the corresponding parameter (core flow in this example).

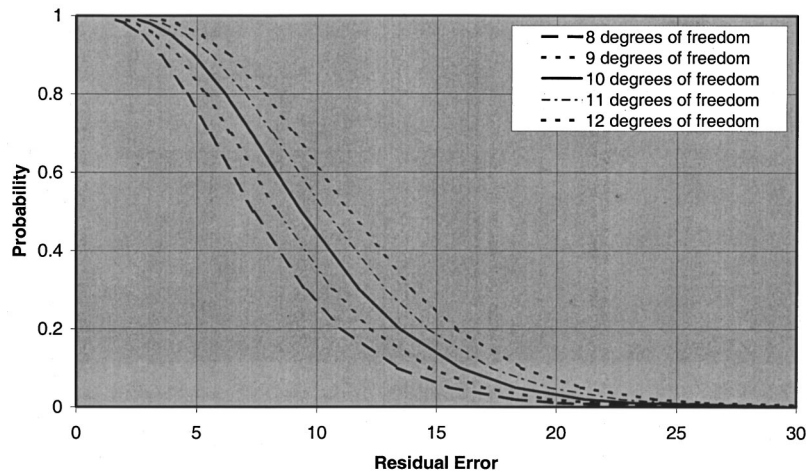


Fig. 4 Chi-squared distribution for 8 to 12 degrees-of-freedom

increase as more independent measurements are added (which conforms to the result that the expected value of $2J_0$ is equal to the number of independent measurements).

This is precisely the interpretation applied for a standard TEMPER solution. When TEMPER invokes the “fault logic,” it changes at least one of the variables so that its standard deviation is no longer unity (by multiplying the standard deviation in the analysis by 100). If the fault being tested is a large error in one of the independent measurements, the behavior is as if that measurement were discarded from the analysis (which would lower the number of degrees-of-freedom by one). The appropriate course of action is less clear if the fault logic is testing a hardware parameter or one of the setting parameters (fan speed, etc.). However, the analysis charges one degree-of-freedom for each variable relaxed in the fault-search analysis. In practice, this means that the number of degrees-of-freedom is reduced by one for most faults. The exception applies to the case where flow and efficiency for a single component are relaxed simultaneously; in this case, TEMPER charges two degrees-of-freedom.

Figure 4 shows the chi-squared distribution, for 8 through 12 degrees-of-freedom. The chart confirms the earlier observation that few engines have a residual error as large as 20 and that a residual error of 10 or less is indicative of a good solution. It also shows that these approximate limits are reduced as the number of degrees-of-freedom (measurements) is reduced. The probability parameter is used because it captures both influences and also because it is more familiar to most users. The fault logic is invoked whenever the base solution has a probability less than 5%. It is deemed to have found a good solution if the faulted solution has a probability greater than 25%.

Using Residual Error and Probability—Practical Considerations

The theoretic discussion of residual error and probability suggests that these parameters may be used to evaluate solutions. Results with high residual error and small probability are suspect, or at least unusual according to the analysis. Unfortunately, these observations rely on two important assumptions.

1. The measurement baselines are good predictions for the engine being tested, in the sense that the average from a large number of engine tests would be equal to the baseline value.
2. The standard deviations are accurate with regard to typical measurement errors and engine-to-engine hardware variations in the sense that they would match the observed standard deviations from a large sample of engines or measurements.

The measurements used in the TEMPER analysis vary from engine model to engine model and also from test cell to test cell. There can be significant differences in the measurement devices, even for the same engine model. For example, some test cells will use a single-element T3 probe where others will use a five-element rake. Several of the measurements are cell-specific (thrust, fuel flow, air flow) including some of the setting parameters (engine inlet pressure and temperature).

There are also reasons to expect differences in the hardware means and standard deviations. Engines are sent to overhaul for various reasons. Some engines receive a complete performance overhaul with work done on every major module. Other engines may be given a core refurbishment or may have only the high pressure turbine reworked. Still other engines receive no performance restoration because the shop visit was prompted by a service bulletin. Thus, the engines being tested include the complete range from fully overhauled modules to modules that have experienced lengthy service with no refurbishment. The mix of engine work scopes varies from one shop to another, providing a source of variation to the hardware statistics. Another lesser cause for variation is that repair practices vary from shop to shop.

The net result of these shop-to-shop differences is that baselines and statistical parameters can be unique for each engine model and test cell (one could even present an argument for further subdivision based on work scope). A sample of at least 20 engines is needed to establish a reasonable baseline for a given engine model and test cell. A larger sample is needed to estimate the statistical parameters used in the analysis (estimates suggest at least 100 distinct engine tests). Getting this number of distinct engine tests is a problem, for shops with heavy work loads; it is impossible for smaller shops.

As a result, the preferred approach for generating the baselines and statistical parameters is seldom used. Baselines are typically developed from the production engine trends, and are adapted to overhaul shops using the cell correlation results. Since cell correlation results are not always available for all parameters, only the prime variables (thrust, fuel flow, EGT, and core speed) are apt to have adjustments. These derived baselines are adjusted using ad hoc methods for estimating typical component performance levels for overhauled engines relative to new production.

TEMPER standard deviations were developed for a few high-volume shops during the initial program development. These were upgraded a few times for other engine models where sufficient data were available or experience showed the existing statistical values to be unsatisfactory. The process for generating the statistical parameters is not rigorous (there are many more unknowns than equations) so that a good deal of judgment is involved in the

specification (Doel [4] gives a brief overview of the process). Many of the more recent TEMPER programs use statistics that were developed for other similar engine models or test cells.

In practice, this means that assumptions 1 and 2 are often not valid. Even though the statistics are not precisely correct from one application to the next, the errors tend to be relatively small and affect one parameter at a time. The more significant problem is with the baselines. The process capability of a cell correlation test introduces errors in the baselines that are larger than if the baselines were derived from a large sample of engines in the overhaul test cell. Not all parameters are included in a cell correlation, so some baselines are derived solely from the production engine trends. Changes may be made in the overhaul test cell measurement system which could alter the baselines. If these changes do not influence the engine shop manual performance calculation, they are unlikely to be communicated; thus, baselines will not be altered. For these and other reasons, there may be sizeable errors in the baselines.

Baseline errors are most easily recognized through monitoring of the residual error and probability. Errors in the baselines inevitably cause the residual error to increase. Small errors are not easily recognized (nor are they important). Larger errors cause the residual errors to increase to values of 40 or more which forces the solution probability to zero. If most analyses give probabilities between 20% and 100%, the baselines are appropriate to the test cell and engine model. An occasional reading with a probability of zero is not unusual. However, if most solutions have probabilities of 1% or less, there is almost certainly a problem in the baselines.

The basic algorithm (excluding the fault logic) is a linear calculation. This means that an error in one of the baselines causes a fixed displacement in the solution for every reading. For example, a baseline error could cause the calculated compressor efficiency to change by 1 point, but this change would occur for every reading processed against the same baseline. An effective way to cope with baseline errors is to perform analyses for all engines and trend the results. The problem engine is recognized because its efficiency or pumping is below the trend established by the other engines. Similarly, a measurement error is recognized when the calculated measurement error lies outside the normal range established by the other engines.

Unfortunately, the fault logic is impacted by the baseline errors. When the residual error is always large, TEMPER will try to find a fault solution. Most of the time it will fail, because no solution can be found that satisfies the requirement for a probability of at least 25%. When it succeeds, there is no assurance that the resulting solution is correct because errors in the baseline can cause the wrong component or sensor error to be selected by the fault logic. The safest way to operate when baseline errors are suspected is to disable the fault logic and use the trend approach described above.

Occasionally, TEMPER will consistently generate low solution probabilities when the baselines are correct. This can happen when one or more sensors have large measurement errors for every reading. In the best of circumstances, TEMPER will recognize a bad measurement and use the fault logic to reject it from the analysis. If another fault is also present, TEMPER cannot simultaneously correct both problems, so it will either generate an incorrect solution (if some third fault can "explain" both large deviations), or more often, it will choose the no fault solution with a large residual error and low probability.

Summary

Most gas turbine diagnostic tools determine component health directly from upstream and downstream measurements or indirectly, using assumptions to fill in for measurements that are not available. These tools are effective when measurement errors are small and when the underlying assumptions are valid. Experienced analysts learn to go beyond a pure reliance on the analysis

tools; they understand the underpinnings well enough to be able to recognize patterns associated with measurement errors or invalid assumptions.

Weighted-least-squares algorithms introduce fundamental changes to the patterns associated with both valid and invalid solutions. They underreport component performance deviations and sensor errors (sensor errors are not estimated by conventional algorithms). True to the idea that the goal of analysis is to support decision making, weighted-least-squares emphasizes correct determination of the faulted component while placing less emphasis on getting the magnitude right.

The algorithms do not rely on a single analysis path; they combine the possible analysis paths statistically, based on the standard deviations for component faults and measurement errors. This behavior contributes to reliability of results for overall engine and core flow. It is less effective for determination of component efficiencies, because redundant measurements are seldom available.

Weighted-least-squares programs are linear, but often employ non-linear extensions to identify large-magnitude faults. TEMPER includes such a feature designated "fault logic." The effectiveness of the fault logic is compromised if the baselines are incorrect. This is often the case for newer engine models where insufficient experience is available. Baseline problems can be detected by examining the solution probability or residual. If the baselines are incorrect, the best strategy is to disable the fault logic and trend results for both good and bad engines. The component faults in problem engines can be recognized because they stand out against the background of normal engine data.

TEMPER is restricted to interpreting the test cell measurements available in the acceptance run. The analyst need not be. A better strategy is to combine the information that TEMPER provides with other information about the engine (on-wing history, prior maintenance, inspection results, etc.). Palmer [7] describes an attempt to automate such an information fusion process using Bayesian belief networks. His work is built upon the efforts of others who were attempting the same synergy manually. A resourceful, experienced user can make TEMPER more effective by exercising good judgment based on the results and on other available information. A better TEMPER will exist when the processes used by such a user are captured and embedded into the program.

Acknowledgments

Many people contributed to the development of TEMPER. Some of the key developers were Dick Henderson, Harlan Durand, Don Berg, Bob Tait, Bernie Kingsolver, and Bob Striebig. Many others contributed to the lore on TEMPER interpretation. Two key contributors in this area were Chuck Humerickhouse and P. (Govin) Govindarajulu.

Nomenclature

- err_{WF} = fuel flow measurement error
- err_{EGT} = exhaust gas temperature (EGT) measurement error
- H = matrix of partial derivatives defining the effect of component faults on measurements (e.g., $\partial WF / \partial \eta_T$)
- J = error term to be minimized by weighted-least-squares analysis
- J_0 = value of J corresponding to "best" weighted-least-squares solution ($2J_0$ is defined to be to the "RESIDUAL ERROR" in TEMPER)
- M = covariance for the component faults (represented by σ_s in the simplified expressions for J in Eqs. (1)–(7)).
- P = percent probability (computed from $2J_0$)
- R = covariance matrix for measurement errors (represented by the σ_m in the simplified expressions for J in Eqs. (1)–(7)).
- RE = residual error ($= 2J_0$)

$\bar{x}$ = vector of component faults (fan flow capacity and efficiency, LPC flow capacity and efficiency, etc.)
 $\hat{x}$ = value of $\bar{x}$ that minimizes J
 $\hat{z}$ = vector of observed measurement deviations (ΔWF , ΔEGT , etc.) from pre-determined baseline
 ΔEGT = observed deviation between measured and predicted EGT
 ΔWF = observed deviation between measured and predicted fuel flow
 $\Delta \eta_F$ = deviation of fan efficiency from baseline
 $\Delta \eta_T$ = deviation of turbine efficiency from baseline
 $\Delta \Gamma_F$ = deviation of fan flow capacity from baseline
 $\sigma_{er_{WF}}$ = standard deviation for fuel flow measurement error
 $\sigma_{er_{EGT}}$ = standard deviation for EGT measurement error
 σ_{Γ_F} = standard deviation for fan flow capacity (overhauled engine-to-engine variation)
 σ_{η_F} = standard deviation for fan efficiency (overhauled engine-to-engine variation)
 σ_{η_T} = standard deviation for turbine efficiency (overhauled engine-to-engine variation)

$\frac{\partial WF}{\partial \eta_T}$ = partial derivative expressing effect of turbine efficiency on engine fuel flow at constant ambient condition and power setting

Superscripts

-1 = denotes the matrix inverse
 T = denotes the transpose of the vector or matrix

References

- [1] Urban, L. A., 1972, "Gas Path Analysis Applied to Turbine Engine Condition Monitoring," AIAA/SAE Paper 72-1082.
- [2] Doel, D. L., 1994, "TEMPER—A Gas-Path Analysis Tool for Commercial Jet Engines," ASME J. Eng. Gas Turbines Power, **116**, pp. 82–89.
- [3] Volponi, A. J., 1982, "A Large Measurement Error Recovery Algorithm for Gas Turbine Module Performance Analysis," Hamilton Standard Report No. 3430.
- [4] Doel, D. L., 1994, "An Assessment of Weighted-Least-Squares Based Gas Path Analysis," ASME J. Eng. Gas Turbines Power, **116**, pp. 366–373.
- [5] Bryson, A. E., Jr., and Ho, Y. C., 1975, *Applied Optimal Control*, Hemisphere, Washington, DC, Chap. 12.
- [6] Zwillinger, D., ed., 1996, *Standard Mathematical Tables and Formulae*, 30th Ed., CRC Press, Boca Raton, FL, Chap. 7.2.
- [7] Palmer, C. A., 1998, "Combining Bayesian Belief Networks with Gas Path Analysis for Test Cell Diagnostics and Overhaul," ASME Paper 98-GT-168.

Setting Up of a Probabilistic Neural Network for Sensor Fault Detection Including Operation With Component Faults

C. Romesis
Research Assistant

K. Mathioudakis
Associate Professor

Laboratory of Thermal Turbomachines,
National Technical University of Athens,
Athens 15710, Greece

The diagnostic ability of probabilistic neural networks (PNN) for detecting sensor faults on gas turbines is examined. The structure and the features of a PNN, for sensor fault detection, are presented. It is shown that with the proposed formulation, a powerful tool for sensor fault identification is produced. A particular feature of the PNN produced is the ability to detect sensor faults even in the presence of engine component malfunction, as well as on deteriorated engines. In such situations, the size of bias that can be identified increases. The way to establish the limits of sensor bias that can be detected is presented along with results from application to test cases with realistic noise magnitudes. The diagnostic procedure proposed here is also supported by an engine performance model. The data used for setting up and testing the PNN are generated by such a model.
[DOI: 10.1115/1.1582493]

Introduction

Detection and isolation of faults that may occur in an engine, at an early stage of their appearance is of great importance to aviation. Measurements of quantities related to engine operation are performed for this purpose. Before they are used for engine condition assessment, it must be ensured that they represent correctly the measured physical quantities, namely that the corresponding sensor readings do not contain a fault. Diagnostic tools allowing sensor fault detection are thus necessary and must be efficient even in cases when component faults or performance deterioration occur.

Sensor fault detection and isolation can be achieved through analytical redundancy methods (Bettocchi et al. [1], Kelly [2], Simani et al. [3], and Bettocchi et al. [4]), using approaches related to the Kalman filter (Caliskan et al. [5]) or fuzzy logic methods (Healy et al. [6]).

Recently, a new type of diagnostic approach using artificial neural networks has become attractive. Among all different neural networks, the back-propagation and the probabilistic neural nets are the architectures, which have mostly been investigated for gas turbine diagnostics. The majority of the researchers refer to performance diagnostics (Eustace et al. [7], Volponi et al. [8], and Pong et al. [9]), while fewer refer to sensor fault detection and isolation. Kanelopoulos et al. [10] studied the performance of back-propagation neural nets for both sensor and actual engine component faults for a single shaft industrial gas turbine. Mattern et al. [11] presented an auto-associative neural network for single sensor fault detection and sensor value estimation on a turbofan engine. Doke et al. [12] used an auto associative neural network for sensor fault detection and isolation of a T56 turboprop engine. Napolitano et al. [13] present a group of back-propagation neural networks for sensor fault diagnosis on military aircraft engines.

Among the advantages of probabilistic neural networks (PNN), the most significant is that training of such a network is known to be much faster than others, such as the back propagation (Bose et al. [14]). Additionally, a PNN performs a probabilistic rather

than a deterministic diagnosis, something which is closer to reality. Comparative and parametric investigations of the diagnostic ability of PNN on turbofan engines have been carried out by Bin et al. [15] and Romesis et al. [16], respectively. Their works have also provided some general information about PNN diagnostic ability.

In the present paper, the use of a PNN for sensor fault detection and isolation of a gas turbine is investigated, by examining implementation into a modern type of civil engine, a high bypass, partially mixed turbofan. Although neural networks have been applied in various other cases, to the authors' knowledge, it is the first time that probabilistic neural networks are being applied for the identification of sensor faults in this type of aircraft engine and such a wide range of sensor fault cases is studied. The network studied in the present work is used for identification of single as well as combined sensor faults, in a "healthy," faulty, and deteriorated engine.

Definition of the Diagnostic Problem

Monitoring the condition of a jet engine is based on a set of values of quantities measured on the engine. Although the procedure we propose is of general validity, it will be easier to present it through its application to a particular engine type. The type chosen is a twin spool, high bypass ratio, partially mixed turbofan engine, representative of those used in civil aviation today. The layout of such an engine is shown in Fig. 1.

A set of measurements for defining operating point and monitoring its condition, typical of engines in service today (Doel [17]), is employed, shown in Table 1. Fuel flow has been chosen as the cycle parameter setting the operating point, a choice made to give worst-case scenarios, since this quantity is linked to significant measurement uncertainty.

Sensor faults are deviations of the readings for these measurements, which are in the form of a bias from the true value. It is the presence of such biases that has to be detected by the diagnostic procedure.

In order to employ a PNN for the detection of sensor faults, a preconditioning of the measurement data is performed: From the measured values, their percent deviation ("delta") from nominal is derived. For a measurement with value Y , coming from the engine, the delta (ΔY) is defined as

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Amsterdam, The Netherlands, June 3–6, 2002; Paper No. 2002-GT-30030. Manuscript received by IGTI, December 2001, final revision, March 2002. Associate Editor: E. Benvenuti.

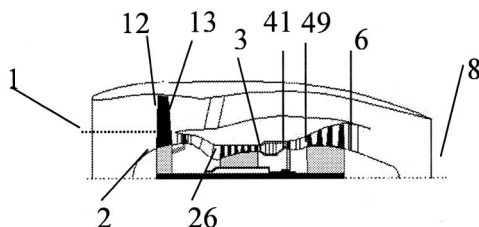


Fig. 1 Layout and station numbering of a turbofan engine

Table 1 Measurements on the engine

(a) Quantities defining operating point (u)		(b) Quantities for monitoring (Y)	
P_1	Inlet total pressure	$XNLP$	LP shaft rpm
T_1	Inlet temperature	$XNHP$	HP shaft rpm
P_{amb}	Ambient pressure	P_{13}	Outer fan outlet pressure
WFE	Fuel flow rate	T_3	HPC outlet temperature
		P_3	HPC outlet pressure
		T_6	LPT outlet temperature
		T_{13}	Outer fan outlet temperature

$$\Delta Y = [(Y - Y_o) / Y_o] \cdot 100. \quad (1)$$

Y_o is the nominal value at the specific operating point (the value we would obtain from a “healthy” engine). It is generated by an engine performance model for the same operating point, defined through an input vector u . The set of deltas thus derived, is fed to the PNN for examining whether sensor faults are present or not.

To ensure that Y_o is a reliable estimation of the nominal value, a model adapted to the particular engine examined must be used. Adaptation ensures that the effect of engine to engine differences or the current engine condition are taken into account and are not thus a source of spurious information (the principles of adaptive

modeling have been discussed by Stamatis et al. [18]). When such a model is used, in the case of an engine with no component or sensor fault, ΔY deviate from zero only because of the presence of noise in the measurements. Component or sensor faults cause additional deviations on the measurements.

In the present work we have employed the engine performance model (EPM) produced by Stamatis et al. [19], adapted to the engine under examination. A feature of this model is that it can simulate engine operation in healthy, faulty and deteriorated condition. Faults and deterioration are represented through appropriate “health parameters.” In the current work, fault parameters represent flow and efficiency factors of different modules of the engine. For a component with entrance at station i of the engine we have

$$\text{Flow factor: } SW_i = (W_i \cdot \sqrt{T_i / p_i}) / (W_i \cdot \sqrt{T_i / p_i})_{ref} \quad (2)$$

$$\text{Efficiency factor: } SE_i = \eta_i / (\eta_i)_{ref}. \quad (3)$$

The use of such factors for describing the health condition of engine components has been discussed by Stamatis et al. [18]. The EPM can thus provide simulated data (the values of the seven measured quantities of the engine) for a given operating point, defined through the four relevant parameters, when the condition of the engine components is known, through the values of the health parameters.

Probabilistic Neural Network—Architecture for Sensor Fault Diagnosis

The PNN architecture proposed here for sensor fault detection is shown in Fig. 2, while its basic principles are given in the Appendix. Inputs to the network are the deltas of the seven quantities measured for monitoring.

Each output node represents a possible sensor fault, while one additional node accounts for all sensors operating with no fault (“Ok” node). Ambient pressure P_{amb} , has not been included,

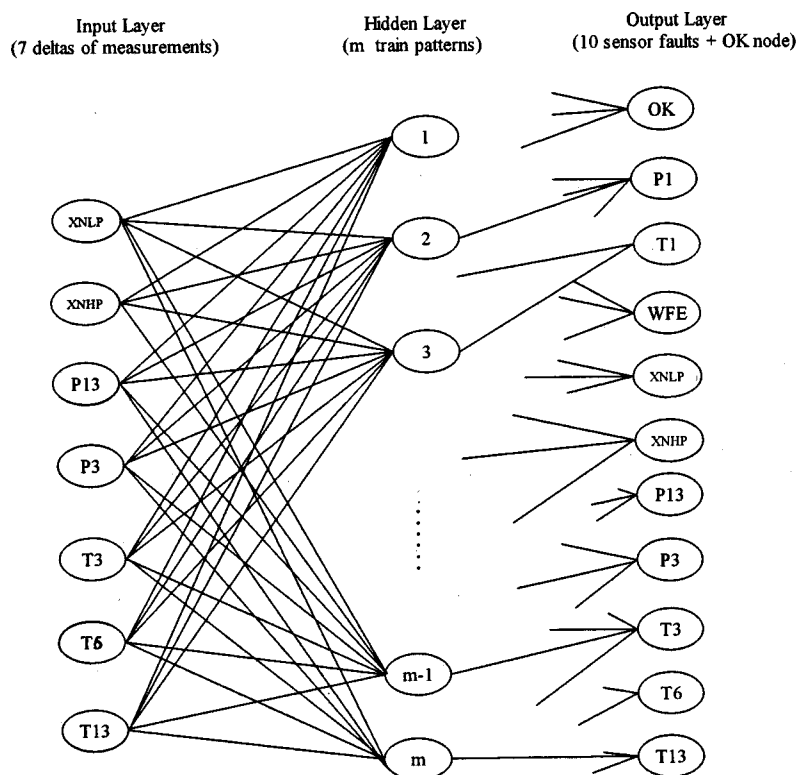


Fig. 2 Probabilistic neural network architecture for turbofan sensor fault diagnosis

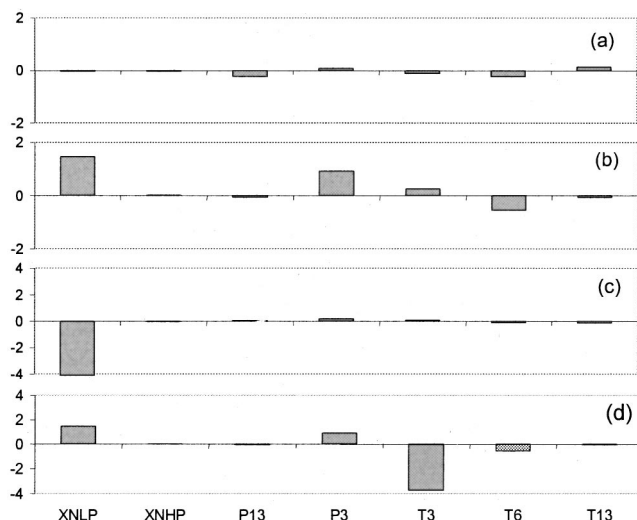


Fig. 3 Patterns of deltas for different situations: (a) healthy engine and sensors, (b) faulty engine, healthy sensor, (c) healthy engine, faulty sensor (XNLP), (d) faulty engine and sensor (XNLP)

since even large deviations of this quantity affect only marginally the engine operating condition. The value produced by the PNN on each node is the probability that the corresponding sensor is faulty.

The intermediate layer contains the training patterns. Each training pattern is a seven-dimensional vector of measurement deltas:

$$x_j = \{\Delta XNLP, \Delta XNHP, \Delta P13, \Delta T3, \Delta P3, \Delta T6, \Delta T13\}. \quad (4)$$

Example patterns are shown in Fig. 3. Under the convention used for this network, patterns (a), (b) are linked to the OK node, while patterns (c), (d) are linked to nodes representing sensor faults.

It must be commented here that deltas imbed information about both the monitoring sensors and the sensors defining the operating condition (the later through the quantity Y_0). This is why the seven-dimensional input vector chosen provides information about the condition of all ten considered sensors.

Generation of Patterns. Patterns of measurement deltas are generated for two purposes: (a) training the network, (b) testing the network. A schematic presentation of the procedure for generating these patterns is shown in Fig. 4. This procedure represents the practical fact, that both monitoring and set-point measurements are known with uncertainty due to the presence of noise.

Noise is added to model predictions, so that values representative of actual measurements are generated. The considered noise levels are presented in Table 2 in the form of percentage of standard deviations from the nominal values at typical cruise conditions, and are those usually encountered for this type of instrumentation (as reported by Curnock [20]).

The different sets generated are shown in Table 3. Training patterns are generated for one operating condition a typical cruise (Cruise-I). This condition is chosen since engine operates for longer intervals in steady state. Additionally, in order to create a more "flexible" network, in the sense of smaller training set and quicker inference, it is desirable to have a network trained with noise-free patterns. Set0 of patterns, including cases of single sensor faults on an engine with several component faults, meets these characteristics and is used for training the network.

In order to examine how the network performs, sets of data covering situations that are expected to be encountered in practice are generated.

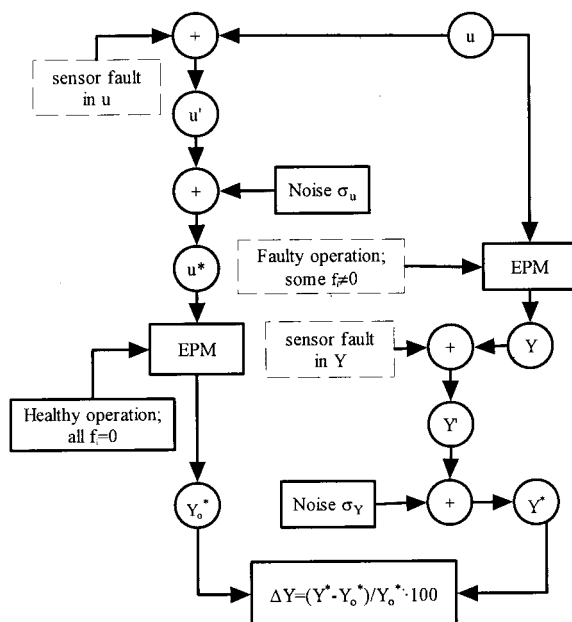


Fig. 4 The procedure for generating the sets of patterns

The first question to be answered is how the network performs when it is presented with patterns containing sensor faults with values not included in the training patterns and in addition are contaminated with typical noise (noise levels of Table 2). Set1 consists of such data. The influence of noise level is then studied. Filtered data are generated, by averaging noisy data samples. One filtered reading is calculated by averaging 20 raw readings, namely values generated using the noise levels of Table 2. The filtered data corresponding to Set1 are contained in Set2. All the subsequent sets are also constituted by using filtered data.

The way of forming the inputs to the network (deltas expressing a percent change) implies that the patterns presented to the network should not be different when the engine operating conditions do not change very much. In order to examine how sensitive the network is to small changes of operating point, Set3 is generated with data from a different operating point at cruise conditions (Cruise-II).

The response of the network to time series of data, for detection of sensor drifts is demonstrated using Set4.

From the definition of a PNN it is expected that it would be suitable for the detection of faults in single sensor readings. In order to examine how it behaves when more than one sensor readings are faulty, Set5 was generated.

As already mentioned, since performance diagnosis is based on measurements, sensor validation must precede any other diagnostic tool. It is, therefore, important for the PNN to maintain a good performance on detecting sensor faults even in the presence of component faults. In order to examine how the network performs in such situations, sensor faults in a faulty engine are included in Set6 and Set7 of patterns.

For the same reason, the PNN should allow efficient sensor validation even in deteriorated engines. This case is represented by Set8 of test patterns.

Examining Probabilistic Neural Networks (PNN) Diagnostic Abilities

The diagnostic performance of the probabilistic neural network described above is tested on the aforementioned sets of test patterns. The procedure followed for inference with the network when a set of measurement values is to be examined, is shown in

Table 2 Noise levels used for pattern generation

σ_u (%)							σ_Y (%)			
$P1$	$T1$	P_{amb}	WFE	$YNLP$	$YNHP$	$P13$	$T3$	$P3$	$T6$	$T13$
0.101	0.027	0.153	0.163	0.051	0.034	0.164	0.094	0.162	0.097	0.230

Fig. 5. This procedure involves the use of the engine performance model to generate the deltas, which are input to the network.

For all the cases presented below, the network is considered to indicate the presence of a sensor fault, when one of the output nodes is assigned a probability greater than 50%.

Effect of Noise Level and Operating Conditions. In order to examine the effect of measurement noise on the performance of the PNN, Set1 and Set2 of patterns are tested. The performance of the network on faults consisting of biases of different magnitude for all individual measurements is demonstrated in Fig. 6(a) and Fig. 6(b), respectively. Performance for biases in the range of $\pm 1\%$ are shown, since larger magnitudes were always detected correctly. Comparison of two figures shows that when data are filtered, the network is capable of detecting deviations of a smaller magnitude. In particular, for unfiltered data, single sensor faults on a healthy operating engine are detectable from a level of deviation equal to 1%, while the same faults in the case of filtered data can be detected even from a magnitude of deviation equal to 0.4%.

The effect of a change in operating point of the engine can be assessed from the results of testing the PNN on two sets representing the same sensor faults at different operating points (Set2 and Set3). The performance of the network for the two cases is demonstrated in Fig. 6(b) and Fig. 6(c), respectively. Comparison of these two figures shows that the network performs equally well for data from operating conditions other than those used for generation of the training data. It is thus insensitive to small changes of operating point.

Establishing Minimum Detectable Faults Levels. It is expected and becomes also apparent from inspection of Fig. 6 that it is easier to detect a sensor fault when it is of a relatively large magnitude. It is desirable to know how large this magnitude needs to be for the fault to be detectable. In order to obtain this magnitude the network is presented with the patterns containing sensor biases of different magnitudes and its performance is tested (Set3 of Table 3). An example of a result for different fault magnitudes of sensor T3 is shown in Fig. 7.

Examining results such as those of Fig. 7, gives the possibility to establish the minimum detectable levels of fault for each individual sensor. These minimum levels are shown in Fig. 8. From this figure it can be concluded that all sensor faults can be detected within a level of $\pm 0.4\%$ to $\pm 0.8\%$. This threshold level is very small, in comparison to the levels reported by methods of other investigators (Simani et al. [3], Bettocchi et al. [4], Volponi et al. [8]) and indicates that the present method is a powerful tool for sensor fault detection.

The minimum detectable levels can be compared to the accuracy of the corresponding measuring instruments, if they are expressed in terms of measurement standard deviation. In Fig. 9 the levels of Fig. 8 are shown, expressed in terms of standard deviation of the corresponding quantities. Since the levels of Fig. 8 do not vary significantly, biases in measurements with higher accuracy seem to be detectable when their magnitude is a larger multiple of their standard deviation, compared to less accurate ones. This is a consequence of the fact that the mechanism of detection

Table 3 The generated sets of patterns

ID	No. of Patterns	Operating Point	Noise in Data	Component Faults Included	Sensor Faults Included
Set0	11352	Cruise-I	No	Each factor: ¹ $\pm(0.5, 1.5, 2.0, 4.0)\%$ and each pair: ² ($a, -a$) & ($-\beta, \beta$) $a = (0.5, 2.5)\%$, $\beta = (1, 4)\%$	(No SF) + Each sensor: ³ $\pm(0.3, 0.7)\%$, $\pm(1.3, 3, 5)\%$, $\pm(10, 40)\%$
Set1	486	Cruise-I	Typical	No	Case A
Set2	486	Cruise-I	Filtered	No	Case A
Set3	486	Cruise-II	Filtered	No	Case A
Set4	400	Cruise-II	Filtered	No	P13: (0–2)% step: 0.01%
Set5	496	Cruise-II	Filtered	No	All combinations of fault ranges: $\pm(2, 3, 4)\%$ for sensors ⁴
Set6	3400	Cruise-II	Filtered	Each factor: ¹ $\pm(1, 3)\%$	Case B
Set7	2000	Cruise-II	Filtered	Each pair of factors: ² ($\pm 3, -1.5\%$) & ($\pm 2, -1\%$)	Case B
Set8	16000	Cruise-II	Filtered	Drifting deterioration of fault parameters ⁵	WFE, YNLP, P3, T6 at ranges: $\pm(0.5, 1, 2, 3)\%$

Notes:

Case A: Each sensor:³ $\pm(4, 4)\%$ per 0.2% and $\pm 7\%$ Case B: Each sensor:³ $\pm(0.5, 1, 2, 3, 4)\%$

¹Considered factors: SW12, SE12, SW2, SE2, SW26, SE26, SW41, SE41, SW49, SE49, A8IMP

²Considered pairs of factors: (SW12, SE12), (SW2, SE2), (SW26, SE26), (SW41, SE41), (SW49, SE49)

³Considered sensors: P1, T1, WFE, YNLP, YNHP, P13, P3, T3, T6, T13

⁴Considered groups of sensors: (YNLP, P13), (P3, T13, T6)

⁵All fault parameters change according to law: $f_i = d_i \cdot I$, where I index representing usage time and the constants d_i express typical deterioration rates

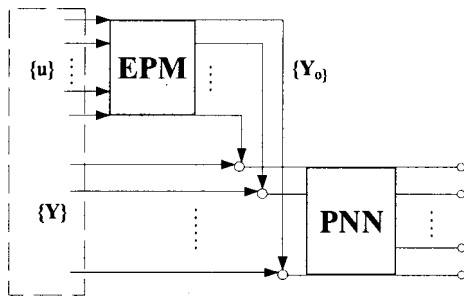
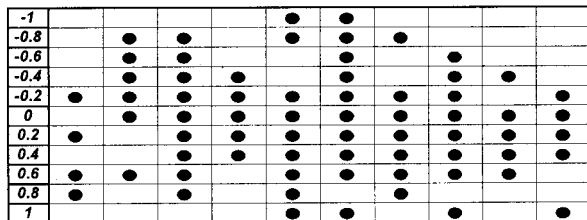
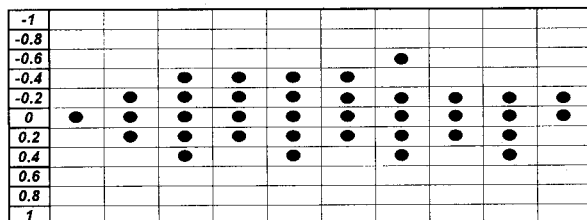


Fig. 5 The procedure for evaluating sensor fault cases

(a) Test patterns of Set1: Trained OP, typical noise



(b) Test patterns of Set2: Trained OP, noise filtered



(c) Test patterns of Set3: Not trained OP, noise filtered

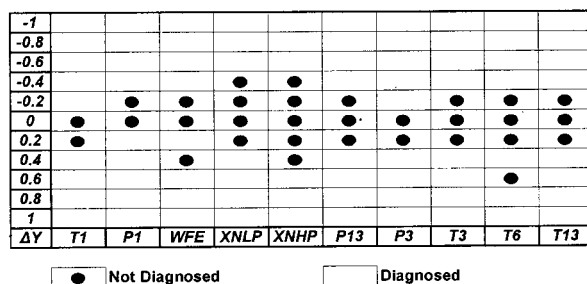


Fig. 6 PNN performance matrix for sensor fault detection

is inherently based on interrelation among all measurements, and thus the behavior of the less accurately measured ones influences the rest.

Detection of Sensor Drifts. The minimum detectable bias is related to how detection is achieved when a sensor reading drifts with time. The way the network responds to such cases is illustrated in Fig. 10 (data from Set4, Table 3). The upper part of this figure (Fig. 10(a)) shows how the fault evolves starting from a not faulty operation, followed by a period that the sensor reading drifts and finally ends to a specific bias.

In Fig. 10(b), the estimated by the PNN probabilities for the described case are shown. It is observed that at a certain moment in time, as the sensor starts drifting from the correct indication,

the probability for it being faulty starts increasing to settle to the value of 100%. Up to that point the probability of all sensor being OK was 100%. The nondisturbed sensors have a zero probability of fault, for the entire interval examined.

Multiple Sensor Faults. Since classes on a PNN are exhaustive and mutually exclusive, the information provided by the PNN structure employed here cannot, in principle, provide information on multiple sensor faults. It has been observed, however, that when the network is presented with multiple sensor fault measurement sets, it always detects a fault on one of the sensors, which are actually faulty. This observation led to a procedure shown schematically in Fig. 11. The idea is that once we know that a fault is present on one sensor, we can eliminate this fault and examine the resulting pattern, in order to find out if an additional fault is indicated. Therefore, after the first pass, the delta of the

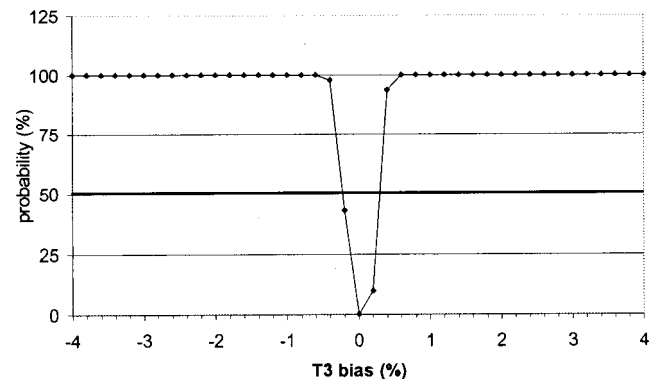


Fig. 7 PNN output for several levels of T3 sensor fault

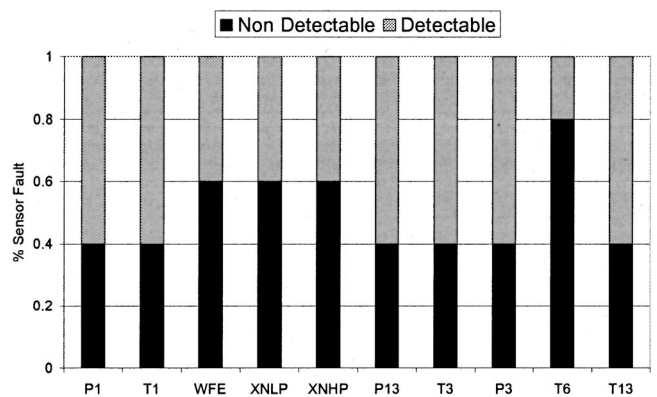


Fig. 8 Minimum detectable levels of sensor faults

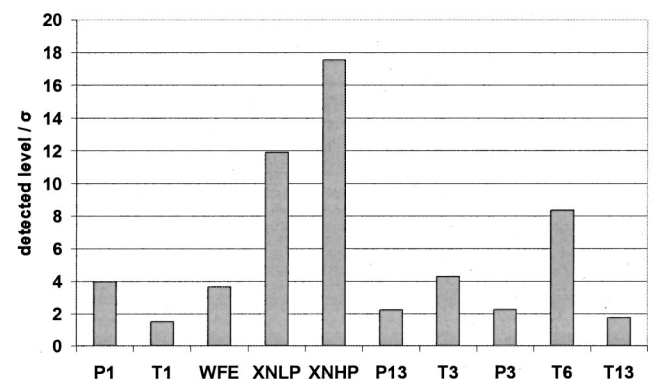


Fig. 9 Minimum detectable levels of sensor faults in terms of measurement standard deviations

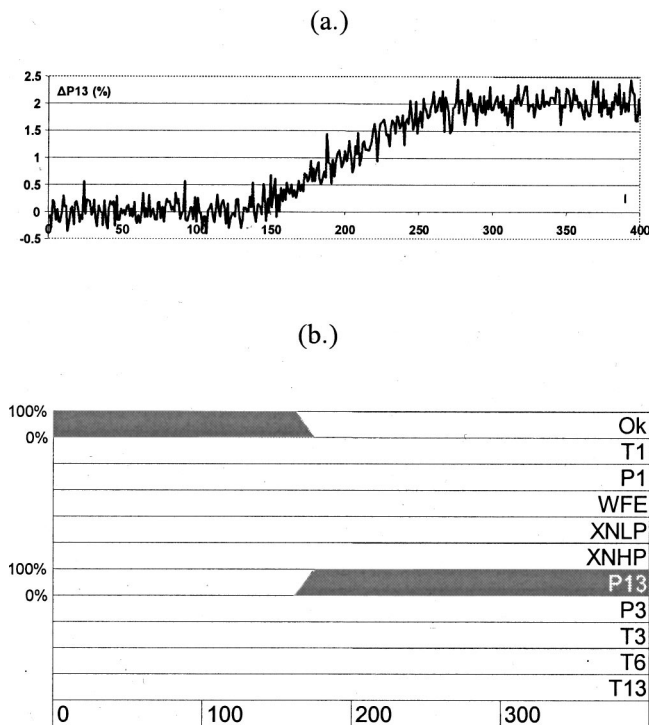


Fig. 10 Detection of a drift in sensor P13: (a) time variation of P13, (b) probabilities produced by the network

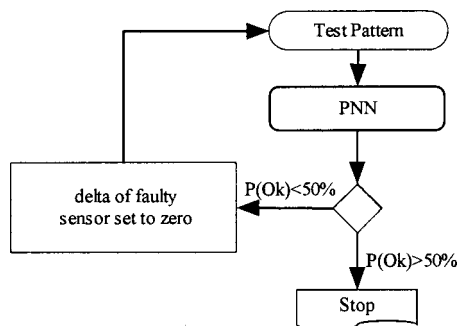


Fig. 11 Procedure followed for multiple sensor fault detection

sensor found faulty is set to zero and the resulting pattern is examined again. A pattern containing a single sensor fault will be found OK from this second pass. If more faults are present, one of them is detected during the second pass. The procedure can continue further, until the pattern is found OK.

This procedure is tested on patterns containing multiple sensor faults (Set5 of Table 3). Two types of multiple sensor faults were considered: faults on two and on three sensors, with several bias magnitudes. It was found that satisfactory performance is achieved for up to three simultaneous sensor faults and that the procedure is effective for faults in one of the seven quantities considered for monitoring. A sample result, indicative of the effectiveness for such cases is shown in Fig. 12. In this figure, the success rate of the PNN is shown (percentage of successfully detected fault cases). The PNN can detect correctly the sensor faults, for a very high percentage of the cases considered.

Sensor Fault Detection in a Faulty Engine. The possibility to detect sensor faults when component faults occur is now examined. The deltas presented to the PNN are generated using the procedure of Fig. 5 in which Y_0 is calculated for a healthy engine. This situation represents the real-life case in which the user has

not yet used the measurements to perform a diagnosis and therefore is not aware if the engine is healthy or faulty.

Component faults are considered as deviations of one single health parameter (single faults) or simultaneous deviation of two health parameters. The presence of such deviations has an effect similar to the effect of the presence of noise, as far as sensor fault detection is concerned. It is therefore expected that the more severe a component fault is (that is greater deviation from the nominal value), and/or the lower the sensor bias, the harder it will be for the PNN to detect correctly the sensor fault.

An example of the performance of the network when tested on data from an engine with faults is shown in Fig. 13. The performance matrix of this figure indicates what levels of XNHP and WFE sensor faults are detectable when two levels of component parameter deviations are considered. The health parameter deviations shown represent individual component faults. The success in detecting a sensor fault is shown for different bias magnitudes, for faults expressed through deviations of each individual engine component health parameters. The particular sensors were chosen as they were found to constitute the best and the worst case, respectively, meaning that detectable levels for other sensors lie between these two cases.

The PNN performance deteriorates when component performance deviations are larger. As in the fault-free engine, it is interesting to know what is the detectable threshold of sensor faults, this time in relation to the magnitude of component fault. An example of minimum detectable levels for XNHP reading, in conjunction with several levels of HP turbine fault expressed through SW41 deviation, is shown in Fig. 14. Although the minimum detectable level seems not to be influenced by small magnitudes of performance deviation (see also Fig. 8), it becomes larger as the component deteriorates further.

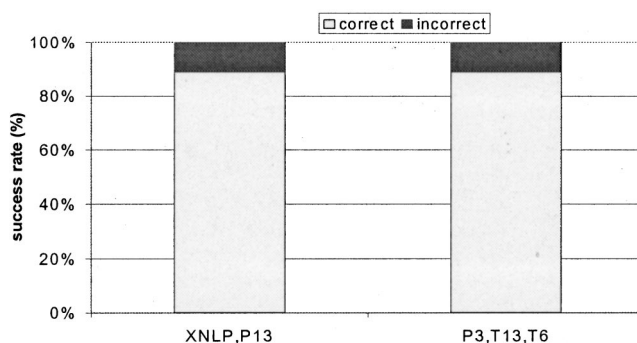


Fig. 12 PNN success rate for cases of multiple sensor faults

		$\Delta XNHP(\%)$ SENSOR FAULT										$\Delta WFE(\%)$ SENSOR FAULT									
		-4	-3	-2	-1	-0.5	0.5	1	2	3	4	-4	-3	-2	-1	-0.5	0.5	1	2	3	4
-3% Deviation	SW12																				
	SE12																				
	SW2																				
	SE2																				
	SW26																				
	SE26																				
	SW41																				
	SE41																				
	SW49																				
	SE49																				
-1% Deviation	ABIMP																				
	SW12																				
	SE12																				
	SW2																				
	SE2																				
	SW26																				
	SE26																				
	SW41																				
	SE41																				
	SW49																				

Fig. 13 Performance of the PNN on detection of sensor's XNHP and WFE fault on a turbofan with single-component fault

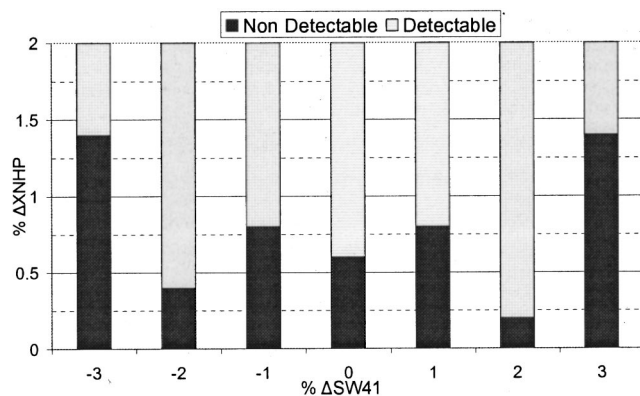


Fig. 14 Minimum detectable levels of $XNHP$ sensor fault, for several levels of HP turbine fault ($SW41$ deviation)

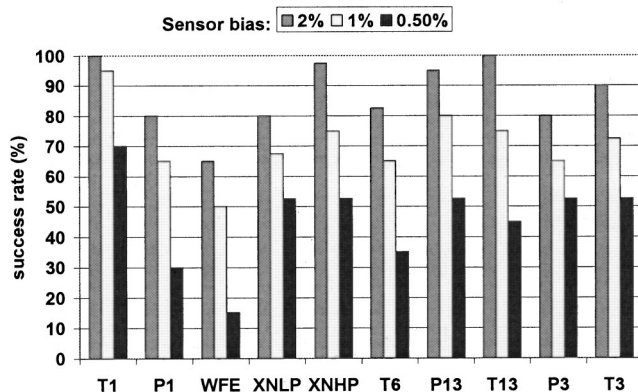


Fig. 15 Performance of the PNN on sensor fault detection on a turbofan with multiple component faults

Finally Fig. 15 gives an indication of the success on detecting three levels of sensor faults (0.5%, 1%, and 2%) for each sensor, in the case of an operation of the engine with multiple component faults translated into a $(\pm 3, -1.5)\%$ and a $(\pm 2, -1)\%$ deviation of one pair of fault parameters each time. The considered pairs of fault parameters are shown in Table 3.

The general conclusion of these figures is that even in the case of the presence of a component fault (single or multiple) in the engine, sensor faults greater than $\pm 1\%$ can be detected efficiently in most cases.

Sensor Faults in a Deteriorating Engine. A situation, which presents an interest to the user, is the possibility for sensor fault detection on a deteriorated engine. In service engine deterioration results in deviations of all component parameters, gradually increasing with time. The situation is similar to the previously examined, when component faults are present, the difference being that all components change simultaneously as engine deteriorates.

When a PNN is used for sensor fault detection, it is of interest to know up to what level of deterioration the performance of the network remains unaltered. Following an approach similar to the previously examined cases, the minimum detectable level of sensor deviations can be established for a given deterioration scenario. Sample results from such a study are shown in Fig. 16. A typical deterioration scenario for a turbofan has been chosen and is represented as a linear change of health indices with usage time of the engine. The x-axis of the figure is a parameter expressing this usage time and could represent flight cycles or some other measure of engine life usage.

In this figure, the limits up to which sensor faults of a given level are detectable are indicated through the vertical lines with

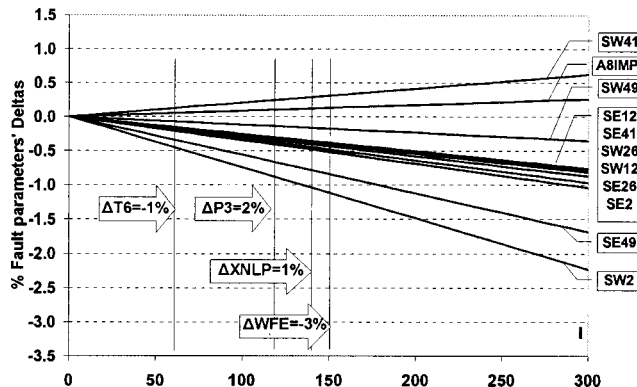


Fig. 16 Performance of the PNN on sensor fault detection on a drifting deteriorated turbofan

the arrow pointing to them. Each line indicates the point up to which the specific sensor faults can be detected by the network trained on healthy engine data. After this point only biases of larger magnitude can be detected. An alternative that would fully restore the network detection capability is the generation of a new training set of patterns, but this time using the EPM adapted to the condition of the engine at that particular instant. This would require that the values of all health parameters are known at that moment, something that can be achieved as for example shown by Mathioudakis et al. [21].

Concluding Remarks

The use of a probabilistic neural network for diagnosis of sensor faults on a turbofan engine has been presented. The fast training and the probabilistic nature of the diagnostic information that such a network provides, makes it a flexible tool for applications.

A particular network type was set up and presented. Its performance was tested on an extensive set of measurements, covering a wide range of practical situations. The diagnostic procedure proposed involves the use of an engine performance model for deriving delta values, which are used by the network for inference.

Testing of the procedure was done in order to establish its capabilities as well as its sensitivity, namely the minimum detectable levels of faults for each type of situation.

The main conclusions concerning the performance of the presented network are as follows:

- The proposed structure of PNN can detect efficiently sensor faults even though the subsequent resulting deviations are of very small magnitude, provided that the engine components are "healthy."
- A simple filtering procedure, like the one considered here, allows efficient diagnosis with PNN trained with simulated noise-free patterns. This leads to more flexible and easy to built networks.
- A network is equally effective for operating points in the vicinity of those used for generation of training information.
- The proposed procedure gives also the possibility to identify multiple sensor faults, including up to three different sensors.
- Sensor faults can be detected even in the presence of component faults or deterioration. The detectable bias levels are then larger. Biases greater than $\pm 1\%$ are detectable in most of such cases.

Acknowledgments

The work for this paper has been carried out in the frame of project BR PR CT 97-0601 and financial support of the Commission of the European Union is gratefully acknowledged.

Appendix

The Principle of the Probabilistic Neural Network. The probabilistic neural network (PNN) is a multilayer feed forward network that classifies input patterns to classes, represented by the output nodes (Bose et al. [14]). When an unknown pattern is presented to the PNN, the network estimates the probability that this pattern belongs to each class.

A schematic presentation of the probabilistic neural network was shown in Fig. 2. The m nodes of the middle layer represent the training patterns χ of the network:

$$x_j = \{a_{1j}, a_{2j}, \dots, a_{nj}\}, j = 1, \dots, m. \quad (5)$$

The k nodes of the third (output) layer represent the classes to which a pattern can be classified into. Each training pattern belongs to one and only one pattern therefore each node X_i of the hidden layer is linked to one and only one output node S_j . For inference, a pattern $x \in \mathcal{R}^m$ is given as an input to the network and an output is produced, which is the probability density functions: $P(S_i|x)$, for each one of the k classes considered. For Gaussian probability density functions we have

$$P(S_i|x) = \frac{P(S_i)}{P(x) \cdot (2\pi)^{m/2} \cdot \sigma_i^m \cdot |S_i|} \cdot \sum_{j=1}^{q_i} \exp \left[-\frac{(x - x_j^{(i)})^T (x - x_j^{(i)})}{2\sigma_i^2} \right] \quad (6)$$

where $x_j^{(i)}$ is the j th pattern of the training set of patterns that “belong” to class i , $|S_i| = q_i$ is the number of the training patterns that “belong” to class i , σ_i is the smoothing parameter of class node S_i , $P(S_i)$ is the a priori probability of class S_i , and $P(x)$ a normalization factor representing the a priori probability of pattern x , which is constant assuming exhaustive and mutually exclusive classes. If we assume equal a priori probabilities for all classes, $P(S_i) = 1/k$, $i = 1, \dots, k$.

Nomenclature

A8IMP = exhaust area
EPM = engine performance model
HPC = high pressure compressor
HPT = high pressure turbine
LPT = low pressure turbine
P_{amb} = ambient pressure
P_i = total pressure at station i
PNN = probabilistic neural network
SE_i = efficiency factor at station i (Eq. (3))
SW_i = flow factor at station i (Eq. (2))
T_i = total temperature at station i
u = vector of quantities defining operating point
WFE = fuel flow rate
W_i = gas mass flow rate at station i
XNHP = high pressure shaft rpm
XNLP = low pressure shaft rpm
Y = vector of measured quantities on an engine
Y₀ = nominal value of quantity Y
*Y** = noisy value of quantity Y

ΔY = percentage deviation of Y (“delta”), Eq. (1)
 σ_Y = vector of standard deviation of vector Y
 σ_u = vector of standard deviation of vector u
 η_i = efficiency of component with entrance at station i

Subscripts

1,2,...,8 = station along the engine gas path, Fig. 1
 ref = the reference values (value in a “healthy” engine)

References

- [1] Bettocchi, R., Spina, P. R., and Azzoni, P. M., 1996, “Fault Detection for Gas Turbine Sensors Using I/O Dynamic Linear Models—Methodology of Fault Code Generation,” ASME Paper 96-TA-2.
- [2] Kelly, R., 1996, “Application of Analytical Redundancy to Detection of Sensor Faults on a Turbofan Engine,” ASME Paper 96-GT-3.
- [3] Simani, S., Spina, P., Beghelli, S., Bettocchi, R., and Fantuzzi, C., 1998, “Fault Detection and Isolation Based on Dynamic Observers Applied to Gas Turbine Control Sensors,” ASME Paper 98-GT-158.
- [4] Bettocchi, R., and Spina, P. R., 1999, “A Method for the Diagnosis of Gas Turbine Sensor Faults in Presence of Measurement Noise,” ASME Paper 99-GT-303.
- [5] Caliskan, F., and Hajiyeve, C., 2000, “Innovation Sequence Application to Aircraft Sensor Fault Detection: Comparison of Checking Covariance Matrix Algorithms,” ISA Trans., **39**, pp. 47–56.
- [6] Healy, T., Kerr, L., and Larkin, L., 1997, “Model Based Fuzzy Logic Sensor Fault Accommodation,” ASME Paper 97-GT-222.
- [7] Eustace, R., and Merrington, G., 1995, “Fault Diagnosis of Fleet Engines Using Neural Networks,” ISABE Paper ISABE 95-7085.
- [8] Volponi, A. J., de Pold, H., Ganguli, Ranjan, and Daguang, Chen, 2000, “The Use of Kalman Filter and Neural Network Methodologies in Gas Turbine Performance Diagnostics: A Comparative Study,” ASME Paper 2000-GT-547.
- [9] Pong, J. L., Ming, C. Z., Tzu, C. H., and Jin, Z., 2000, “An Evaluation of Engine Faults Diagnostics Using Artificial Neural Networks,” ASME Paper 2000-GT-29.
- [10] Kanelopoulos, K., Stamatis, A., and Mathioudakis, K., 1997, “Incorporating Neural Networks Into Gas Turbine Performance Diagnostics,” ASME Paper 97-GT-35.
- [11] Mattern, D., Guo, T., Graham, R., and McCoy, W., 1998, “Using Neural Networks for Sensor Validation,” prepared for the 34th Joint Propulsion Conference Paper No. NASA/TM-1998-208483.
- [12] Doke, R., and Singh, R., 1999, “Neural Networks for the Detection of Gas Turbine Sensor Faults,” ISABE paper ISABE 99-7255.
- [13] Napolitano, M., An, Y., and Seanor, B., 2000, “A Fault Tolerant Flight Control System for Sensor and Actuator Failures Using Neural Networks,” *Aircraft Design 3*, Pergamon, New York, pp. 103–128.
- [14] Bose, N. K., and Liang, P., 1996, *Neural Network Fundamentals With Graphs, Algorithms and Applications*, McGraw-Hill, New York.
- [15] Bin, S., Jin, Z., and Shaoji, Z., 2000, “An Investigation of Artificial Neural Network (ANN) in Quantitative Fault Diagnosis for Turbofan Engine,” ASME Paper 2000-GT-32.
- [16] Romessis, C., Stamatis, A., and Mathioudakis, K., 2001, “A Parametric Investigation of the Diagnostic Ability of Probabilistic Neural Networks on Turbofan Engines,” ASME Paper 2001-GT-11.
- [17] Doel, D., 1992, “TEMPER—A Gas Path Analysis Tool for Commercial Jet Engines,” ASME Paper 92-GT-315.
- [18] Stamatis, A., Mathioudakis, K., and Papailiou, K. D., 1990, “Adaptive Simulation of Gas Turbine Performance,” ASME J. Eng. Gas Turbines Power, **112**, pp. 168–175.
- [19] Stamatis, A., Mathioudakis, K., Ruiz, J., and Curnock, B., 2001, “Real Time Engine Model Implementation for Adaptive Control & Performance Monitoring of Large Civil Turbofans,” ASME Paper 2001-GT-362.
- [20] Curnock, B., 2001, “OBIDICOTE Program Work Package 4 Steady-State Test Cases for Engine Deterioration,” Rolls Royce Report, Document Number DNS78608, May.
- [21] Mathioudakis, K., Kamboukos, Ph., and Stamatis, A., 2002, “Turbofan Performance Deterioration Tracking Using Non-Linear Models And Optimization Techniques,” ASME Paper GT-2002-30026.

Nonlinear Engine Component Fault Diagnosis From a Limited Number of Measurements Using a Combinatorial Approach

N. Aretakis
Research Associate

K. Mathioudakis
Associate Professor

A. Stamatis
Research Associate

Laboratory of Thermal Turbomachines,
National Technical University of Athens,
Iroon Polytechniou 9,
Athens 15773, Greece

A method for diagnosing component faults of jet engines is presented. It uses nonlinear gas path analysis techniques to determine the values of health parameters, with the help of a suitably formulated engine simulation model. The incentive of the method is to achieve the determination of the values of component health indices when a limited number of measured quantities is available, which do not allow the determination of all the fault indices simultaneously. A combinatorial approach is introduced, in order to circumvent the problem of the insufficient information for determining a full set of indices. After obtaining a set of possible solutions, a selection procedure is applied to isolate the ones that can give the actual fault identity. Quantification of the fault comes at a final step, when the faulty component has been identified. Different scenarios of faults on a twin spool partially mixed turbofan engine are considered in order to demonstrate the effectiveness of the method. The limitations of the method are also discussed.

[DOI: 10.1115/1.1582494]

Introduction

The multiple benefits offered by implementation of diagnostic methods into engine monitoring schemes are well understood today. Description and analysis of those benefits can be found in many published works in recent years.

Monitoring and diagnostic systems used commercially by jet engine manufactures employ the Kalman-type formulation of linear gas path analysis, together with some “augmentation,” to improve its effectiveness (Urban and Volponi, [1], Barwell [2], Provost [3], and Doel [4,5]). Doel [5] has given a critical review of such techniques. One of the shortcomings he identified was the inability of such algorithms to actually estimate the magnitude of true deviation caused by a component fault, usually underestimating the real problem and attributing parts of it to other component faults. These techniques need to use some a priori information about engine condition. The sensitivity of the techniques to an a priori estimate of engine condition was also commented by Volponi [6], (“it works well when you already know the answer”).

The fundamental reason of this behavior lies in the fact that only a limited number of performance related quantities can be measured on an engine. A component condition diagnosis on the other hand, requires the determination of a usually larger number of unknown component “health” parameters. The fact that measurement readings can be faulty themselves and the need for determination of sensor faults, increases further the number of unknowns (for a description of the problem, see, for example, Urban and Volponi [1] and Doel [4]).

A method to overcome the problem of fewer measurements than parameters has been introduced by the research group of the present authors, (gas path analysis at discrete operating points Stamatis et al. [7]). The idea is that instead of installing additional sensors on an engine, data from existing sensors at several discrete operating points can be exploited, to derive a larger number

of parameters. The potential of this method has been appreciated by different research groups and its implementation in different ways has been the subject of recent studies (Kobayashi and Simon [8], Groenstedt [9], and Gulati et al. [10]). This technique, however, at least in the ways it has been implemented up to day, is efficient when operating points are clearly discrete, namely they are far apart in the engine-operating envelope. Inefficiencies for closely spaced operating points were observed by Groenstedt [9], while Gulati et al. [10] also identified similar inadequacies in their implementation.

A method to overcome these problems is proposed in the present paper. It uses data from one operating condition of the engine and derives information about faults related to components characterized by “health” parameters larger in number than the available measurements, without sacrificing the accuracy of component deviations magnitude. Applicability of the method is demonstrated on an extensive set of fault cases, while its limitations are also discussed.

Background, Some Basics of Diagnostics

For the purpose of the present analysis an engine is considered as a system, whose operating point is defined by means of a set of variables, denoted as $\mathbf{u}$. The health condition of its components is assumed to be represented through the values of a set of appropriate “health” parameters, contained in a vector $\mathbf{f}$. The system is observed through measured variables (speeds, pressures, temperatures, etc), contained in a vector $\mathbf{Y}$. The operating engine establishes a relationship between these parameters, which can be expressed through a functional relation:

$$\mathbf{Y} = \mathbf{F}(\mathbf{u}, \mathbf{f}). \quad (1)$$

For a given operating point $\mathbf{u}$ the measurement values depend only on the health condition of engine components.

$$\mathbf{Y} = \mathbf{F}(\mathbf{f}) \quad (2)$$

Measured quantities are thus produced by such a relation, from given values of health parameters. The problem of diagnostics is to seek a solution to the inverse problem, namely to determine the values of the health parameters from a given set of measurements.

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Amsterdam, The Netherlands, June 3–6, 2002; Paper No. 2002-GT-30031. Manuscript received by IGTI, December 2001, final revision, March 2002. Associate Editor: E. Benvenuti.

Occurrence of particular faults can then be detected if deviations of health parameters from the reference state are observed.

In linear gas path analysis, the health parameters are represented as the unknown “deltas” of component performance parameters (typically efficiency and mass flow capacity). They are related to known measurement “deltas” through relations produced by linearization of the general nonlinear relations, assuming small deviations. They are then treated as a linear system of equations, to be solved for determining the unknowns (e.g., Urban and Volponi [1]). In nonlinear methods, the full equations are treated directly with no simplification. An example of such a method, the method of adaptive modeling (Stamatis et al. [11]), uses “modification factors” as health parameters and solves for them through an optimization procedure.

For either type of method, in order to be able to derive values of parameters without the need for additional a priori information, the number of equations must be equal to or larger than the number of unknowns. In such a case it is in general possible to invert relation (2) and produce the unknown parameter values:

$$\mathbf{f} = \mathbf{F}^{-1}(\mathbf{Y}). \quad (3)$$

Implementation of a procedure materializing this relation in practical situations presents a difficulty because:

- measured values always contain noise.
- the number of measuring instruments is limited. Information available at a specific operating condition may not be sufficient for the derivation of all parameters (fewer equations than unknowns).

The presence of noise is dealt with by processing a number of readings, and appropriate statistical techniques. The fact that measurements are fewer than parameters necessitates the addition of equations. A minimum norm of the parameters vector deviations is the approach usually adopted as, for example, in Kalman-type methods. This means that even if noise-free measurements were available, the solution produced will not necessarily be the precise one, although it may be close to it. The alternative approach of discrete operating conditions may provide a solution, but it is also subjected to limitations, as mentioned in the Introduction.

The method we present here can overcome this problem, for the particular case of component fault identification, as explained in the following section.

The Principle of the Method

The approach we follow is based on the fact that the problem to solve is a physical problem and not just a mathematical one, for the solution of a system of equations. This means that when we formulate the algorithm for the solution of the problem, we should take into account all existing information and not just a part of it. Usually in diagnostic methods the part of the information exploited is what is contained in the equations representing the engine model. The minimum norm addition is an arbitrary assumption, which has been adopted mainly because of the empirical fact that it seems to work well in practice. The physical situation is usually included in what Doel [5], calls augmentation techniques (“concentrator” for COMPASS, “Fault Logic” for TEMPER). In order to understand the reasoning behind our procedure, we first have to comment on three types of situations that are encountered in practice, when an identification problem has to be solved for assessing the condition of an engine. The three types are:

1 Engine Deterioration: The performance of all individual engine components degrades with time. In order to identify the condition of a deteriorated engine the health parameters of all components need to be determined. However, knowledge of the way components deteriorate from previous experience may provide a number of additional relations, which can help the solution of the diagnostic problem.

2 Engine Component Faults: Changes in one or more components occur, which lead to degradation of their performance. In

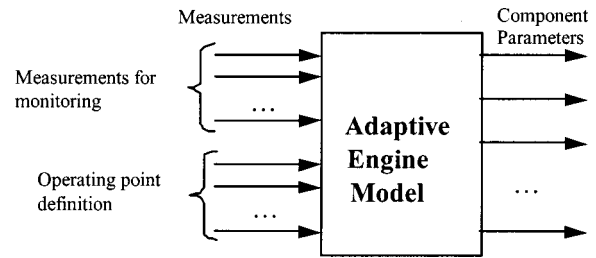


Fig. 1 Derivation of component condition parameters using an adaptive engine model

such situation it is known that some but not all components have sufficiently changed. This information may be exploited for solving the diagnostic problem.

3 Individual Engine Identification: The problem to be solved is the determination of the performances of all individual components of a particular engine, when the performance of an average engine is known. The performance of each component can deviate from the average one in either direction (mainly due to manufacturing and assembly tolerances). In this case all component parameters have to be identified and there is no a priori knowledge as to which of them deviate and in what direction. This is the most difficult case for identification.

In the present paper we will deal with problem 2, component fault detection and identification. We take advantage of the following facts:

- It is not probable that all engine components will suffer a fault simultaneously. A fault may affect one or some of the engine components. This observation comes from existing experience.
- When we have a number of measurements available, it is possible to determine the values of health parameters of a faulty component, provided that we formulate an appropriate system of equations. If we are able by some means to reduce the set of candidate faulty components to a number less than or equal to the number of measurements, then their values can be determined. The question, however, is, how can we identify this reduced set of candidate components, out of the full one.

The procedure we propose here allows the choice of appropriate sets, by examining all possible combinations of parameters, obtaining solutions and establishing a criterion for isolating the faulty component. The tool that we will use for determining health parameters, from an equal number of measurements, will be the technique of adaptive modeling, which is fully nonlinear. When a set of measurement data is available and fed to the model, a set of component condition parameters is derived as shown in Fig. 1. Description and different cases of implementation of this technique into different types of gas turbines have been presented by the present authors (e.g., Stamatis et al. [11], Tsalavoutas et al. [12], and Mathioudakis et al. [13]).

In order to facilitate understanding, the method will be presented through its implementation on a modern type jet engine.

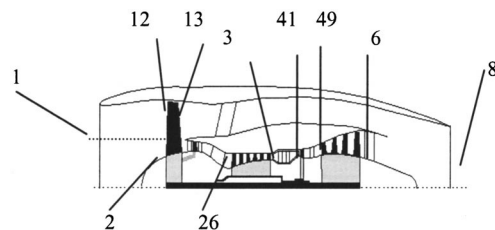


Fig. 2 Layout of a turbofan engine and station numbering for positions of interest

Table 1 Measurement used for monitoring

	Measurement	Symbol
1	LP shaft Rpm	XNLP
2	HP shaft Rpm	XNHP
3	Fan outer pressure	P_{13}
4	HP compressor outlet pressure	P_3
5	HP compressor outlet temperature	T_3
6	LP turbine outlet temperature	T_6
7	Fan outer temperature	T_{13}

Table 2 Component condition parameters

Component	Health Parameter	Symbol	No.
Outer fan	Flow factor	SW_{12}	1
	Efficiency factor	SE_{12}	2
Fan inner	Flow factor	SW_2	3
	Efficiency factor	SE_2	4
HPC	Flow factor	SW_{26}	5
	Efficiency factor	SE_{26}	6
HPT	Flow factor	SW_{41}	7
	Efficiency factor	SE_{41}	8
LPT	Flow factor	SW_{49}	9
	Efficiency factor	SE_{49}	10
Nozzle	Exhaust area	A_8	11

Engine Layout and Representation

A partially mixed, high bypass ratio turbofan, representative of the type used in civil aviation today, is used as a test case. One reason for this choice is the usefulness of this type of application for today's aviation. Another reason is that this type of engine is one of the most complex ones used in this field today. It constitutes therefore a "hard" test case for the method proposed. In Fig. 2 the layout of the test engine, with the measurement locations numbered, is presented.

The operating point of the engine is defined through four quantities (components of $\mathbf{u}$): ambient pressure (P_{amb}), total inlet temperature (T_1) and pressure (P_1), and the fuel flow rate (WFE). The set of measurements used for monitoring is shown in Table 1. It consists of a typical set used for monitoring such engines performance (for typical measurements sets see, for example, Doel [4]).

Parameters employed to characterize the condition of each component are a flow factor SW and an efficiency factor SE. They determine how component flow capacity and efficiency respectively is modified from reference condition. For a component with entrance at station i of the engine, we define

$$\text{Flow factor: } SW_i = \frac{W_i \cdot \sqrt{T_i}}{P_i} \bigg/ \left(\frac{W_i \cdot \sqrt{T_i}}{P_i} \right)_{\text{ref}}, \quad (4a)$$

$$\text{Efficiency factor: } SE_i = \frac{\eta_i}{\eta_{i_{\text{ref}}}}. \quad (4b)$$

Therefore the components of vector $\mathbf{f}$ are the factors for all major engine components, shown in Table 2. The use of such factors for describing the health condition of engine components has been discussed by Stamatis et al. [11].

For the purposes of the present work the values of quantities that would be measured on the engine are produced by an engine performance model (EPM) (Stamatis et al. [14]). The EPM provides simulated data (that is the values of the seven measured quantities of the engine) for a given operating point, defined through the four relevant parameters, when a set of values of the 11 condition parameters, is known. Although the EPM is not referred to a specific commercial turbofan engine in service, its structure reflects fully the principles and the technology of modern turbofan engines in civil aviation. Data thus produced can be contaminated with noise, to resemble actual measured data.

An adaptive model has been built on the basis of the EPM,

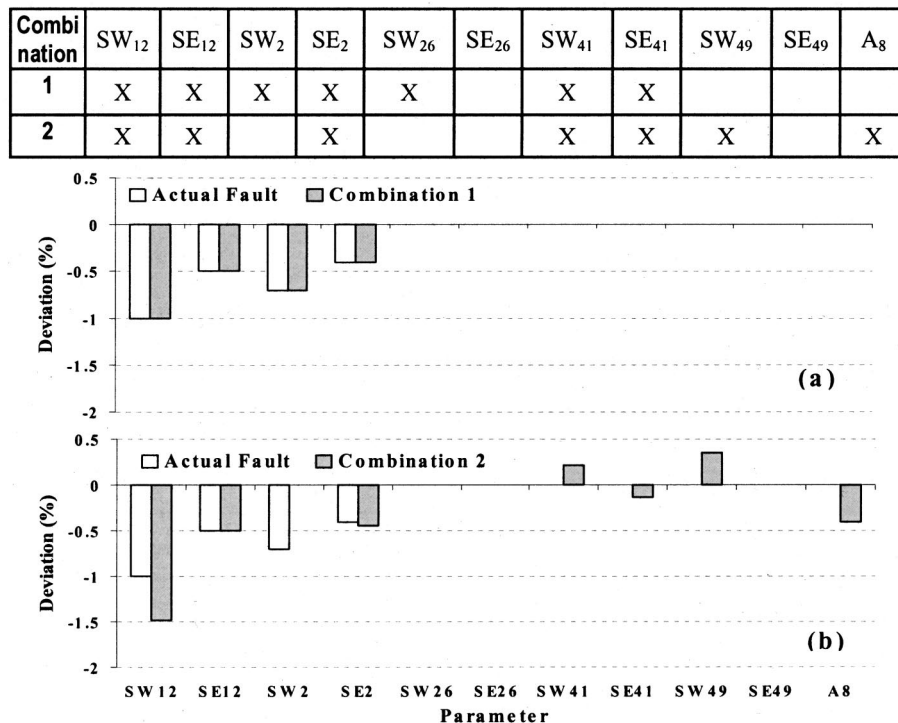


Fig. 3 Health parameter deviations using a combination that (a) contains all actual fault parameters, (b) doesn't contain all actual fault parameters

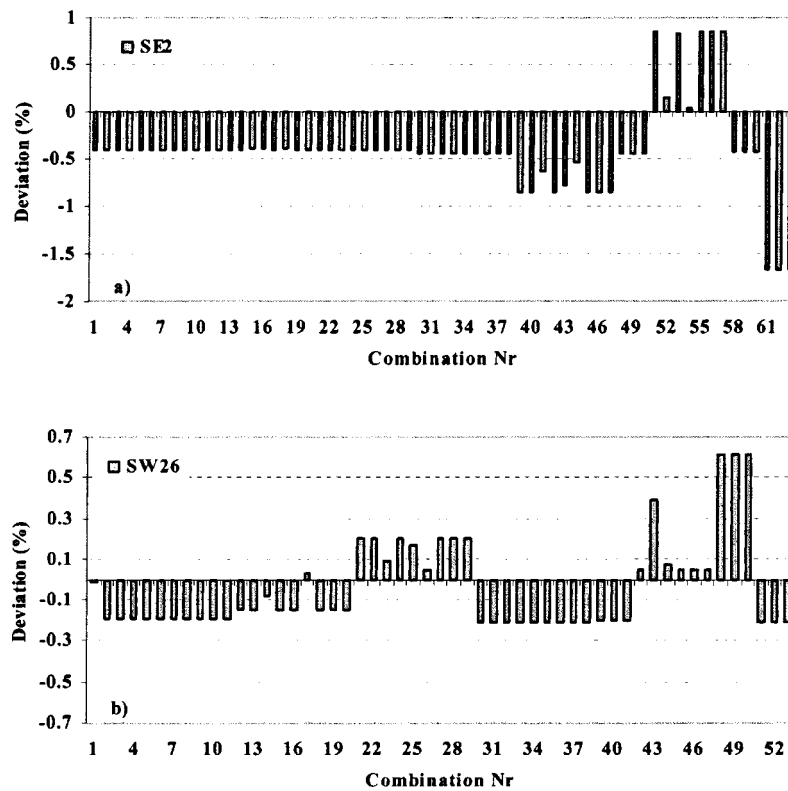


Fig. 4 Deviations of estimated fault parameter values using different combinations, for parameter (a) SE₂ (actually affected), (b) SW₂₆ (not affected)

which is used as a diagnostic tool and can produce seven out of the 11 condition parameters when the seven measurements are known.

The Combinatorial Diagnostic Approach

We describe now, how the values of component health parameters can be estimated, when a component fault is present. First, we observe that it is not possible to calculate all parameters, because we have only seven measurements available and therefore seven equations. Since the number of measurements is less than the number of health parameters, an infinite number of sets of 11 parameter values satisfying the seven equations is possible.

If the components infected by a fault were a priori known, then it would be possible to choose a smaller number of health parameters as unknowns, and derive their values (provided that this number is not larger than seven). If a number of seven parameters is chosen, then a 7×7 nonlinear system is solved. If the number is less than seven, then the solution will be achieved through a nonlinear least-squares approach, guaranteeing a best fit to the measurement data.

Given that faulty components are not a priori known, a procedure for identifying them has been devised as follows: The measurement set is fed to the adaptive diagnostic model and a unique solution (seven parameters out of 11) is produced. The procedure

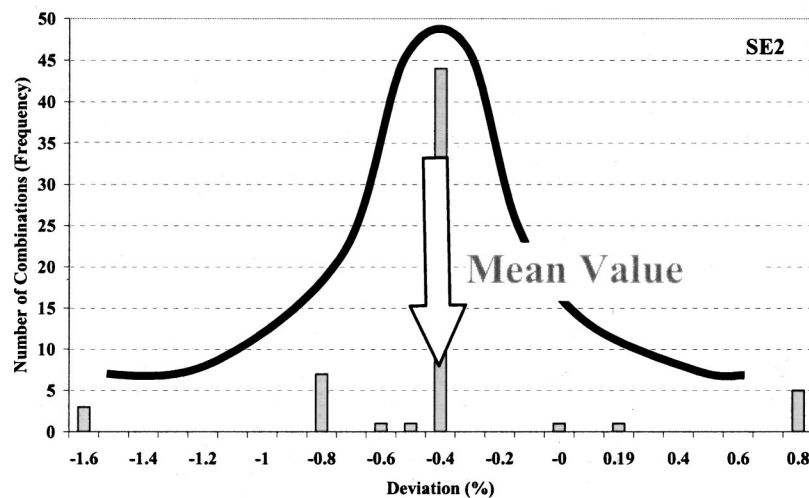


Fig. 5 Parameter distribution calculation for SE₂, data of Fig. 4(a)

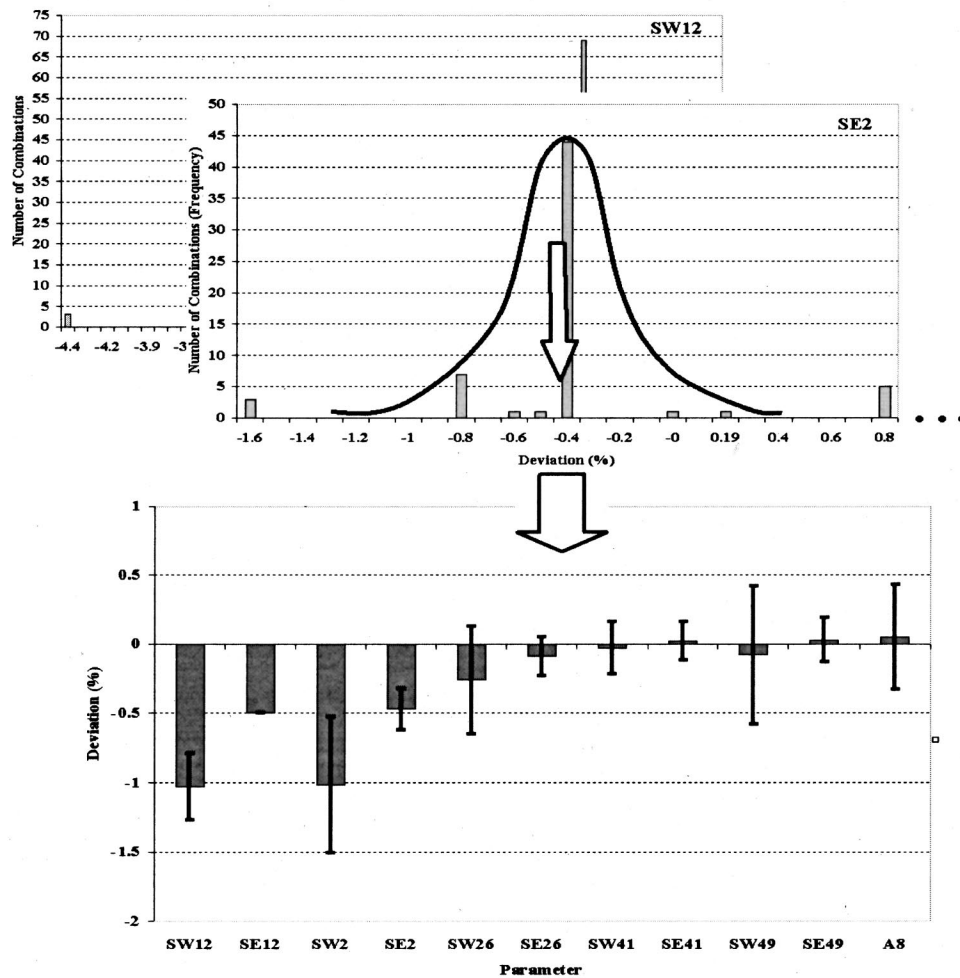


Fig. 6 The way of producing the most probable fault signature

is applied for all different possible parameter combinations, namely as many times as the number of all possible combinations of 11 per seven:

$$\text{Number of combinations} = \binom{11}{7} = 330.$$

For each different parameter combination, which may or may not contain the examined fault, a different set of parameters deviations is derived (we will term the set of such deviations a “fault signature” hereafter).

The idea that if an appropriate combination of unknowns is

Table 3 Percent deviations of health indices values for simulating faults

Fault ID	Outer Fan		Fan Inner		HPC		HPT		LPT		Nozzle
	SW12	SE12	SW2	SE2	SW26	SE26	SW41	SE41	SW49	SE49	A8
A	-1	-0.5	-0.7	-0.4							
B		-1									
C					-1	-0.7					
D						-1					
E					-1						
F							1				
G							-1	-1			
H								-1			
I										-1	
J									-1	-0.4	
K									-1		
L									1	-0.6	
M											1
N											-1
O											2

Table 4 Measurement noise

Sensor	Pamb	P1	T1	WFE	XNLP	XNHP	P ₁₃	P ₃	T ₃	T ₆	T ₁₃
3 sigma	100 Pa	100 Pa	2 K	2 gr/s	6 rpm	12 rpm	300 Pa	5 KPa	2 K	2 K	2 K

chosen, then a diagnosis may be achieved was first mentioned by Urban [15]. He did not provide a procedure for isolating such combinations but rather suggested a “subjective” choice after inspection of different sets, on the basis of existing experience, as to which modules may be more susceptible to faults.

Let's consider now the example of measurement data set coming from a fan fault, described through the following health parameter deviations: $SW_{12} = -1\%$, $SE_{12} = -0.5\%$, $SW_2 = -0.7\%$, $SE_2 = -0.4\%$. The data are fed to the adaptive diagnosis model and a solution for a set of unknown parameters including these ones is derived. In this case the solution reproduces precisely the values of these health parameters, as shown in Fig. 3(a) (noise-free values have been used at this point, just in order to demonstrate the principles). If, on the other hand, the set of unknowns considered doesn't contain all four health parameters that have produced the data, the solution contains deviations of other parameters. Take as an example the case of a combination not containing SW_2 , which means that SW_2 is assumed to not deviate from its reference value. The solution contains then deviations on other health parameters, as shown in Fig. 3(b). It is nevertheless observed that, although SW_2 is not contained in the set, the remaining three deviations are predominant in the estimated signature and not far from their actual values.

It should be noted here that it is not possible, in general, to derive a solution for all possible combinations of measurements and component parameters, since the system of equations can be ill conditioned for some of them. Therefore, only the combinations with good observability are used. Observability is examined on the basis of the condition number of the linearized form of the system of equations, as described in detail by Kamboukos et al. [16].

Derivation of solution for all selected parameter combinations produces a set of values for each one of the 11 parameters. Examples of two such sets, obtained for parameters SE_2 and SW_{26} in the aforementioned case, are shown in Fig. 4. SE_2 , being a parameter actually affected, exhibits estimated values mostly close to the actual deviation value, Fig. 4(a). SW_{26} values, on the other hand, exhibit a scattering around zero deviation, Fig. 4(b).

This observation provides a suggestion, that if estimations are averaged, it may be possible to derive values close to the actual ones for each parameter. Estimated parameter values are thus treated as if they were measurement data, for derivation of mean values and standard deviations. A procedure described below is used to set criteria for assessing which components are affected.

From the available estimated values for each parameter, the parameter value distribution is calculated. A normal distribution is then fitted to these data, using the mean and the standard deviation calculated over the entire set of values. The result for the sample parameter SE_2 used above is shown in Fig. 5.

Subsequently, a new estimate of the mean value and standard deviation is calculated by rejecting all points lying outside a pre-defined range around the mean value. For the test cases presented here the range of 95% confidence level has been used.

This procedure is applied for each individual health parameter. The results are then combined to produce a fault signature, which contains parameter deviation values, indicative of the actual deviations. An example of such a signature is shown in Fig. 6.

Using these results a decision on which component has suffered a fault is taken on the basis of a diagnostic index DI. First a diagnostic index is derived for each one of the component parameters

$$DI_{x_i} = \frac{|\bar{x}_i|}{\sigma_{x_i}} \quad (5)$$

where $\bar{x}_i$ is the mean value and σ_{x_i} the standard deviation for a parameter $\bar{x}_i$. Then a diagnostic index is calculated for the component according to the following equation:

$$Diag_i = \sqrt{DI_{1i}^2 + DI_{2i}^2} \quad (6)$$

(In the case of nozzle that has only one parameter we take $DI_{2i} = 0$).

A faulty component is identified as the one that corresponds to the *diagnostic index of maximum value*.

In the results presented below, the outcome of this procedure will be designated detection phase or as *1st pass*.

Estimation of Deviations. The first pass has produced values for deviation that are suitable for indicating faulty components. It is possible to improve the actual magnitude of estimated deviations, taking advantage of the fact that the faulty component has now been detected.

The procedure described above is repeated, but this time only combinations containing the health parameters of the component found faulty are taken into account. This second screening of the solutions is designated as *2nd pass*. It is used for two purposes: (a) as a verification of the result obtained through the 1st pass, (b) to produce deviation magnitudes, which are closer to the actual values.

The final stage, that produces the best estimation of deviation magnitudes, is the application of a best-fit procedure for derivation of the magnitude of the deviations of the parameters of the component found faulty. Adaptive modeling is applied, formulated as an optimization procedure, so that the measured values are approximated as closely as possible. The problem solved is the calculation of the health parameter values of the faulty component, so that the following function is minimized:

$$OF = \sum_{i=1}^M \left[\frac{1}{\sigma_{y_i}} \frac{y_i^{\text{model}} - y_i^{\text{measured}}}{Y_i^{\text{measured}}} \right]^2 \quad (7)$$

This calculation provides estimated parameter deviations with the minimum uncertainty allowed by the available set of measurements.

Table 5 Fault detection results for examined cases

Fault	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Noise-free data	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1
Noisy data	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1

1: correct detection 0: incorrect detection

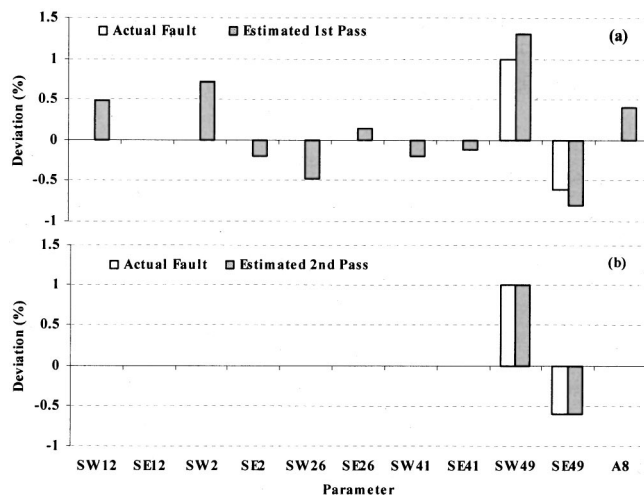


Fig. 7 Estimated fault signatures: (a) first pass, (b) second pass, noise-free data, fault L

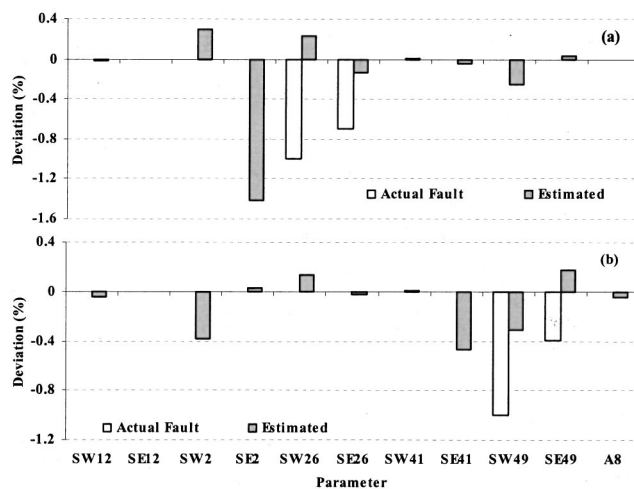


Fig. 8 Estimated fault signatures: (a) fault C, (b) fault J, second pass, noise-free data

Application Test Cases

In order to evaluate the effectiveness of the proposed method, test runs were performed on data produced for a number of cases covering a representative range of jet engine component faults. Table 3 shows the fault cases considered, which cover possible faults in all individual components. This set of cases has been defined in the frame of a European Research project, and has been reported by Curnock [17].

Two types of data sets were used: (a) noise-free data. The reason for this choice is that we want to examine how the method performs for situations where an exact solution exists. It will reveal the method abilities related on the principle used and not due to lack of information because of the presence of noise. (b) data to which noise was added. This test is more realistic from an application point of view, since actual data will always be containing noise. The noise levels employed are shown in Table 4.

In order to make the examined test cases more close to real engine operation where during a cruise flight small variations of operating conditions are faced, a sequence of data sets for different operating conditions and for each one of the examined faults was generated. This sequence contains 50 flight conditions, which are typical of readings taken every five minutes through the cruise sections of a flight. A period of cruise at an intermediate altitude is followed by a longer period at the main cruising altitude. Small variations in flight speed, altitude, and outside air temperature are included.

For all the examined cases the method was applied in the following way: Since a faulty situation is described from a number of data sets (sequence), the method is applied to each individual set, as described in the previous section, with the difference that averaging is performed on the values derived from all sets in the sequence.

Application Test Results. The overall identification results for both noise free and noisy data sets are shown in Table 5.

Observation of the results shows that the proposed method allows the correct identification of a significant percentage of the faults provided as a benchmark. The faults, which seem to be nonidentifiable, are fault C and fault J, for which there is an observability problem with the current measurement set as will be proved in the following. The correct detection of fault C and the failure in the case of fault K for noisy data sets is due to information counterfeiting because of the presence of noise.

In order to show the kind of result derived for the estimated values of health parameters, typical estimation results for both noise-free and noisy data sets are shown in Figs. 7–11. The esti-

mated fault signatures compared with the actual faults are shown in these figures.

In Fig. 7 a comparison between 1st pass and 2nd pass estimated values for fault L and noise free data is shown. It's observed that, although there is not a difference between 1st and 2nd pass concerning the detection possibilities, 2nd pass improves significantly the estimated values.

In Fig. 8 the estimated fault signatures for the two fault cases (Fault C and J) that cannot be detected, for noise-free data are presented.

We observe that the method for both cases gives as more possible infected component the one close to the truly infected. For example for fault C gives fan inner instead of HPC, while in the case of fault J gives HPT instead of LPT. A reason for this behavior may be that the current measurement set does not include measurements between inner fan and HPC and between the two turbines. This fact causes the method to treat in both cases the two close components as one, leading to the results of Fig. 8.

Similar observations arise from noisy data sets results. Indicative results are shown in Figs. 9 and 10. Actual and estimated fault signatures as well as the corresponding diagnostic index are shown in these figures.

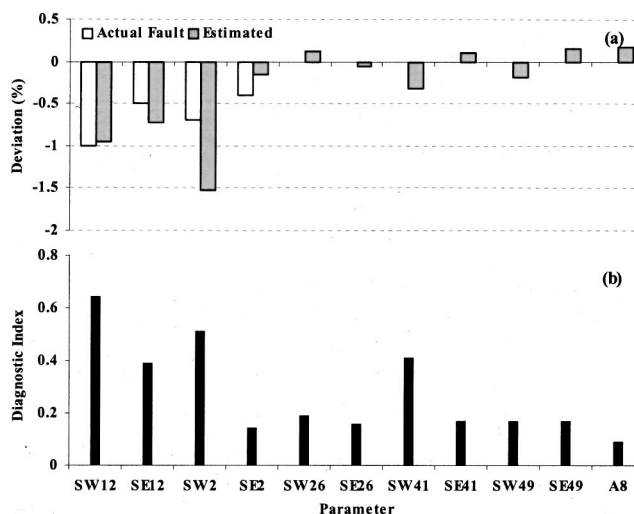


Fig. 9 (a) Estimated fault signature and (b) corresponding diagnostic index, second pass, noisy data, fault A

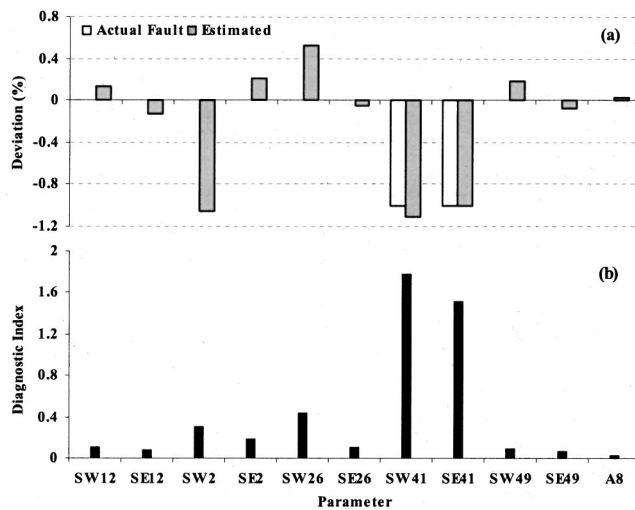


Fig. 10 (a) Estimated fault signature and (b) corresponding diagnostic index, second pass, noisy data, fault G

The first one corresponds to a very clearly detected case (fault A), in the sense that parameter deviations are larger for the faulty component. The second one also corresponds to a correctly detected case (fault G), but some deviation (SW_2) rises to a non-negligible size. The diagnostic index still provides a clear indication of the faulty component.

The estimation of deviation values can be further improved using the best fit procedure described above, after the 2nd pass as shown for example in Fig. 11 for the case of fault A and noisy data sets.

The estimated deviation values produced with best fit procedure for all examined cases are presented in Table 6, compared to the

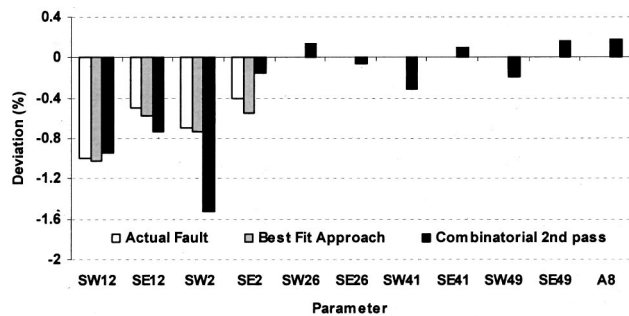


Fig. 11 Comparison between second pass and best fit procedure estimation results for fault A and noisy data

Table 6 Estimation of deviation values through adaptation to measurements

Fault					Fault				
A	Act.	SW12: -1	SE12: -0.5	SW2: -0.7	SE2: -0.4	I	Act.	SW49: 0	SE49: -1
	Est.	-1	-0.6	-0.7	-0.6		Est.	0	-1.1
B	Act.	SE12: -1	SW2: 0			L	Act.	SW49: 1	SE49: -0.6
	Est.	-1.3	-0.1				Est.	1	-0.7
D	Act.	SW26: 0	SE26: -1			M	Act.	A8: 1	
	Est.	0.2	-1.1				Est.	1	
E	Act.	SW26: -1	SE26: 0			N	Act.	A8: -1	
	Est.	-1	-0.1				Est.	-1	
F	Act.	SW41: 1	SE41: 0			O	Act.	A8: 2	
	Est.	1	0				Est.	2	
G	Act.	SW41: -1	SE41: -1						
	Est.	-1	-1.1						
H	Act.	SW41: 0	SE41: -1						
	Est.	0	-1.1						

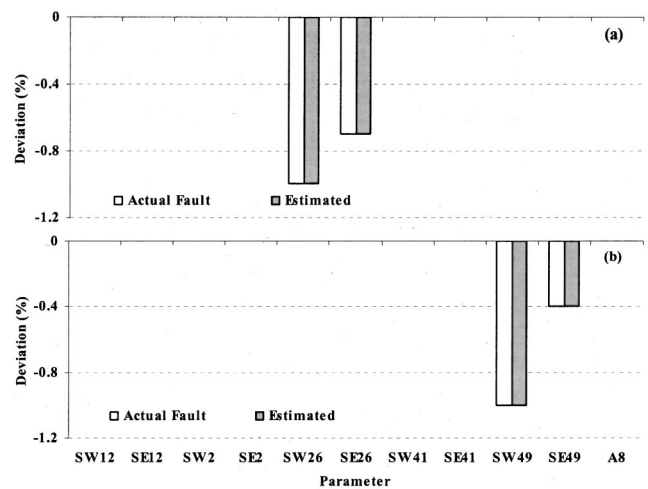


Fig. 12 Estimated fault signatures in the case of using additional measurements: (a) fault C, (b) fault J, second pass, noise-free data

actual ones. It's observed that the estimated values are very close to the actual ones.

Discussion

The effectiveness of the method has been demonstrated for a particular set of measured quantities. The inability to isolate and identify correctly faults affecting both efficiency and flow capacity in the high pressure compressor and low pressure turbine (faults C and J), is related to this particular measurement set. This can be demonstrated by results obtained with the addition of measured quantities.

In order to detect fault C the additional measurement must be between fan inner and HPC. For that reason, the measurement chosen was the pressure at HPC inlet. The estimated fault signature for fault C using additionally this measurement is shown in Fig. 12(a).

We observe that the estimated fault signature is identical with the actual fault. However, use of this additional measurement still does not allow detection of fault J. This fault can be detected if a measurement between the two turbines is also added. The measurement proven to be the best choice for that purpose was the pressure at LPT inlet. The estimated fault signature adding both measurements for fault J is shown in Fig. 12(b) which is also similar to the actual one. The conclusion drawn from this examination is that the present method can distinguish faults affecting these components, if appropriate information between the components is available.

It must be pointed out that extended data sets are usual in engine test cells, but they may also be possible on wing. For example, Doel [4] reports that interturbine measurements can be done on wing as well.

It was also shown that the presence of noise may lead to some small confusion, but this is related to only the fact that the magnitude of deviations considered here is rather small with respect to the level of noise in the measurements. It was verified that the situation becomes clear, resembling the case of noise-free data, when larger deviation magnitudes are present.

A feature of the present method is that it does not require any a priori knowledge of the condition of components. In this respect it presents an advantage over Kalman type methods, in that it faces situations with fewer measurements than parameters, but it still can provide accurate estimation of the actual deviation magnitude. In this respect it presents a similarity to the "concentrator" approach employed by Barwell [2].

The way it is implemented suggests that the method is mostly suitable for detection and identification of situations with component faults. It must be commented here that application has been attempted to the case of deterioration. In this case, there was no selection of parameters and 2nd pass, but averages were derived over all the solutions produced by combinations. The result was that the overall picture and proportions of deviations were captured, but the accuracy of the values obtained was not very good.

Although methods for solving the diagnostic problem when parameters are fewer than the measurements exist (e.g., Stamatis et al. [7]) the method presented offers an advantage: It can be applied to data coming from one operating point or in any case neighboring operating conditions, for which discrete operating point gas path analysis cannot be employed. It can be employed when data from operation only at cruise condition are available. On the other hand, the method implementation is rather fast, since it involves solutions of exact systems of equations, while the minimization problem solved is within a space with small dimension. These features make it possible for on-board application, while it presents a powerful tool for test cell data analysis as well.

Conclusions

A method for diagnosing faults in the components of a jet engine was presented, suitable for cases that a limited instrumentation set is available.

The method is applicable to situations where data are available at steady operating conditions and can both isolate and identify the fault in individual components.

The method was tested on an extensive set of component faults, with an instrumentation set typical for today's engines. It was found that it can detect and identify correctly faults in all components, with some limitation for some HP compressor and LP turbine faults. It was shown that this weakness is eliminated with the addition of inter-spool measurements, while keeping the measurement set realistic.

Acknowledgments

The work for this paper has been carried out in the frame of project BR PR CT 97-0601 and financial support of the Commission of the European Union is gratefully acknowledged.

Nomenclature

- A_8 = exhaust area
- DI = parameter diagnostic index (Eq. (5))
- Diag = component diagnostic index (Eq. (6))
- EPM = engine performance model
- $F(\cdot)$ = function representing engine model

- $\mathbf{f}$ = health parameters vector
- HPC = high pressure compressor
- HPT = high pressure turbine
- LPT = low pressure turbine
- OF = Objective function (Eq. (7))
- P_{amb} = ambient pressure
- P_i = total pressure at the station i of the engine
- SE_i = efficiency factor of component with entrance at station i (Eq. (4))
- SW_i = flow factor at position i of the engine (Eq. (4))
- T_i = total temperature at the position station i
- $\mathbf{u}$ = vector of quantities defining operating point
- WFE = fuel flow rate
- W_i = air mass flow rate at the station i
- $\bar{x}_i$ = health parameter mean value
- XNHP = high pressure shaft rpm
- XNLP = low pressure shaft rpm
- $\mathbf{Y}$ = vector of measured quantities of an engine
- η_i = efficiency of component with entrance at station i
- σ_Y = standard deviation of Y

Sub and Superscripts

- 1,2,...,8 = station along the engine gas path, Fig. 1
- ref = reference value, value in "healthy" engine

References

- [1] Urban, L. A., and Volponi, A. J., 1992, "Mathematical Methods of Relative Engine Performance Diagnostics," SAE Trans. **101**, J. Aerosp. Sect. 1, Technical Paper 922048.
- [2] Barwell, M. J., 1987, "Compass-Ground Based Engine Monitoring Program for General Applications," SAE Technical Paper 871734.
- [3] Provost, M. J., 1988, "COMPASS: A Generalized Ground Based Monitoring System," Paper 42 AGARD-CP-448, Engine Condition Monitoring, Technology and Experience.
- [4] Doel, D., 1994, "TEMPER—A Gas Path Analysis Tool for Commercial Jet Engines," ASME J. Eng. Gas Turbines Power, **116**, pp. 82–89.
- [5] Doel, D., 1994, "An Assessment of Weighted-Least-Squares Based Gas Path Analysis," ASME J. Eng. Gas Turbines Power, **116**, pp. 366–373.
- [6] Volponi, A., "Sensor Error Compensation in Engine Performance Diagnostics," ASME Paper 94-GT-58.
- [7] Stamatis, A., Mathioudakis, K., Berios, G., and Papailiou, K., 1991, "Jet Engine Fault Detection With Discrete Operating Points Gas Path Analysis," J. Propul., **7**(6), pp. 1043–1048.
- [8] Kobayashi, T., and Simon, D., 2001, "A Hybrid Neural-Genetic Algorithm Technique for Aircraft Engine Performance Diagnostics," 37th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit, July 8–11, Salt Lake City, UT, Paper AIAA-2001-3763.
- [9] Groenstedt, T. V., 2001, "A Multi Point Gas Path Analysis Tool for Gas Turbine Engines With a Moderate Level of Instrumentation," XV ISABE, Bangalore, India, Sept. 3–7, Paper ISABE-2001-1139.
- [10] Gulati, A., Taylor, D., and Sign, R., 2001, "Multiple Operating Point Analysis Using Genetic Algorithm Optimization for Gas Turbine Diagnostic," XV ISABE, Bangalore, India, Sept. 3–7, paper ISABE-2001-1139.
- [11] Stamatis, A., Mathioudakis, K., and Papailiou, K. D., 1990, "Adaptive Simulation of Gas Turbine Performance," ASME J. Eng. Gas Turbines Power, **112**, pp. 168–175.
- [12] Tsalavoutas, A., Mathioudakis, K., Stamatis, A., and Smith, M., 2000, "Identifying Faults in the Variable Geometry System of a Gas Turbine Compressor," ASME Paper 2000-GT-33.
- [13] Mathioudakis, K., Stamatis, A., Tsalavoutas, A., and Aretakis, N., 2001, "Performance Analysis of Industrial Gas Turbines for Engine Condition Monitoring," Proc. Inst. Mech. Eng., Part A, J. Power Energy, **215**(A2), pp. 173–184.
- [14] Stamatis, A., Mathioudakis, K., Ruiz, J., and Curnock, B., 2001, "Real Time Engine Model Implementation for Adaptive Control & Performance Monitoring of Large Civil Turbofans," ASME Paper 2001-GT-0362.
- [15] Urban, L. A., 1975, "Parameter Selection for Multiple Fault Diagnostics of Gas Turbine Engines," AGARD CP-165, Diagnostic and Engine Condition Monitoring, June Paper 19.
- [16] Kamboukos, Ph., Oikonomou, P., Stamatis, A., and Mathioudakis, K., "Optimizing Diagnostic Effectiveness of Mixed Turbofans by Means of Adaptive Modelling and Choice of Appropriate Monitoring Parameters," AVT Symposium on "Monitoring and Management of Gas Turbine Fleets for Extended Life and Reduced Costs," Manchester, UK, 11 Oct. 8–11.
- [17] Curnock, B., 2000, "OBIDICOTE Project—Work Package 4. Steady-State Test Cases," Rolls-Royce Report DNS62433.

Degradation Effects on Combined Cycle Power Plant Performance— Part I: Gas Turbine Cycle Component Degradation Effects

A. Zwebek

e-mail: gotross@hotmail.com

P. Pilidis

e-mail: p.pilidis@cranfield.ac.uk

Department of Power Engineering and
Propulsion,
School of Engineering,
Cranfield University,
Bedford MK43, 0AL, UK

This paper describes the effects of degradation of the main gas path components of the gas turbine topping cycle on the combined cycle gas turbine (CCGT) plant performance. First, the component degradation effects on the gas turbine performance as an independent unit are examined. It is then shown how this degradation is reflected on a steam turbine plant of the CCGT and on the complete combined cycle plant. TURBOMATCH, the gas turbine performance code of Cranfield University, was used to predict the effects of degraded gas path components of the gas turbine have on its performance as a whole plant. To simulate the steam (bottoming) cycle, another Fortran code was developed. Both codes were used together to form a complete software system that can predict the CCGT plant design point, off-design, and deteriorated (due to component degradation) performances. The results show that the overall output is very sensitive to many types of degradation, especially in the turbine of the gas turbine. Also shown is the effect on gas turbine exhaust conditions and how this affects the steam cycle.

[DOI: 10.1115/1.1519271]

Introduction

Combined-cycle power plants achieve very good economic performance. This economic performance is dependent on several factors; prominent among them are first cost, high efficiency, low maintenance cost, and good availability.

Good availability is partly influenced by downtime. It is possible to reduce downtime by having detailed knowledge prior to any inspection of the faulty components of a gas turbine. It is, therefore, very important to know how the CCGT power station behaves if different components of the plant are degraded. It is important to establish how various faults manifest themselves in terms of variations of measurable parameters such as temperatures and pressures. Furthermore, a more detailed knowledge of the engine will allow the user to take some of the maintenance action when it is necessary. This will also help reduce operating costs.

This paper describes how common faults, affect CCGT plant performance. In this paper, the degradation of the gas turbine components is examined, while the other bottoming plant (steam turbine plant) were kept at its original design point conditions. Much has been written on the effects of gas turbine degradation, but there seems to be a scarcity of material on the effects of degradation on combined cycle performance.

Gas Turbine Performance Deterioration

The main gas path components of the gas turbine, namely compressor and turbine, will degrade with engine use, [1–5], which then results into engine performance deterioration.

The focus of this paper is the simple-single shaft gas turbine coupled with a single-pressure HRSG bottoming cycle, of which a schematic diagram shown in Fig. 1. More complex steam cycles are currently in service, but the results of the present work will be a useful guide.

The gas turbine of this unfired cycle is an engine of conventional design. This engine has the following (design point) specifications:

inlet mass flow	= 408.66 kg/sec
compressor pressure ratio	= 15.2
turbine entry temperature	= 1697.80 K
exhaust mass flow	= 419.4 kg/sec
exhaust temperature	= 871.24 K
power	= 165.93 MW
Thermal efficiency	= 35.57%

In order to cover the widest range of faults that may occur in the gas turbine, it was assumed that each component may degrade separately, and then all components were assumed to degrade together, to establish the nature of the faults and to assess if they are additive or not. The faults investigated were the following:

1. compressor fouling,
2. compressor erosion,
3. compressor isentropic efficiency degradation,
4. turbine fouling,
5. turbine erosion,
6. turbine isentropic efficiency degradation,
7. compressor and turbine fouling,
8. compressor and turbine erosion, and
9. compressor and turbine isentropic efficiency degradation.

For a clearer and complete discussion of the terms (fouling and erosion) mentioned above and for their mechanism, the reader may wish to examine other work, [1,2].

Because the combustion system is not likely to be a direct cause of gas turbine performance deterioration, [1], it was assumed not to degrade for the following reasons:

1. The faults in combustion chamber that affect overall performance are rare in comparison to those faults that may occur in the compressor and turbine.
2. Any malfunction in the combustion chamber would mean increased emissions, which is not allowed by environmental laws in many places.

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, New Orleans, LA, June 4–7, 2001; Paper 2001-GT-0388. Manuscript received by IGTI, December 2000, final revision, March 2001. Associate Editor: R. Natile.

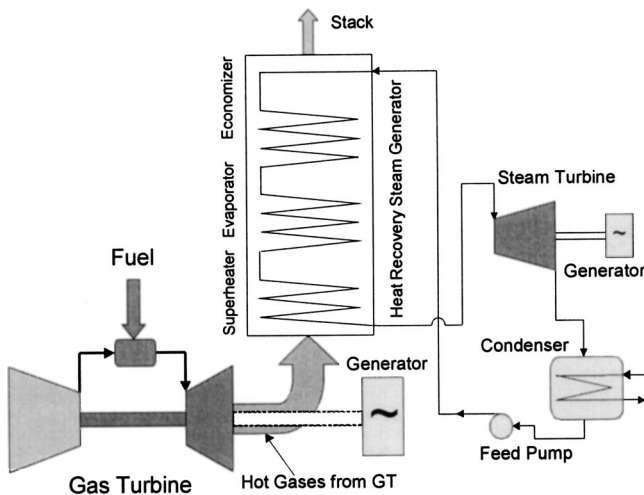


Fig. 1 Schematic diagram of a single pressure CCGT power plant

Faults Representation

In the current study, in order to predict the gas turbine degradation effects, faults implanted on each component of the gas turbine are defined in the following way:

Fouling. Compressor or turbine fouling is represented by reduced flow capacity at the inlet of the component plus a reduction in the component isentropic efficiency. By doing so, it is assumed that there is a blockage in the inlet area of the component due to fouling along with a decrease in the component isentropic efficiency due to surface roughness, for example.

Erosion. Compressor erosion is represented by a lower inlet mass flow capacity and a reduction in compressor isentropic efficiency. On the other hand, turbine erosion is represented by an increased flow capacity plus a reduction in the turbine isentropic efficiency, [2].

These two phenomena are represented by changing the so-called nondimensional mass flow Eq. (1) of the component maps (Table 1).

$$\frac{\dot{W}\sqrt{T_i}}{PA} = \text{constant} \quad (1)$$

Component Efficiency Degradation. This is modeled by reducing the component isentropic efficiency of the appropriate map and keeping all other parameters at their design point (DP) levels. In this case, it was assumed that the component isentropic efficiency may decrease from its DP value due to any reason, such as blade tip rubs or FOD.

To assist the reader observing these faults in a more clearer and readable way, these faults are summarized in Table 1.

Table 1 Representation of component degradation

Fault	Represented by	Range
Compressor fouling	drop in Γ	0.0–(–5.0%)
Compressor erosion	drop in η_C	0.0–(–2.5%)
Turbine fouling	drop in Γ	0.0–(–5.0%)
Turbine erosion	drop in η_T	0.0–(–2.5%)
FOD	drop in η_T	0.0–(–2.5%)
	drop in η_C and η_T	0.0–(–5.0%)

Table 2 Component isentropic efficiency variation with degradation

Physical Fault	Nondimensional Mass Flow Change (A)	Isentropic Efficiency Change (B)	Ratio A:B
Compressor fouling	$\Gamma_C \downarrow$	$\eta_C \downarrow$	~1:0.5
Compressor erosion	$\Gamma_C \uparrow$	$\eta_C \downarrow$	~1:0.5
Compressor corrosion	$\Gamma_C \downarrow$	$\eta_C \downarrow$	~1:0.5
Turbine fouling	$\Gamma_T \downarrow$	$\eta_T \downarrow$	~1:0.5
Turbine erosion	$\Gamma_T \uparrow$	$\eta_T \downarrow$	~1:0.5
Turbine corrosion	$\Gamma_T \downarrow$	$\eta_T \downarrow$	~1:0.5
Foreign object damage	$\Gamma_{C/T} \downarrow$	$\eta_{C/T} \downarrow$	~1:2.0
Thermal distortion	$\Gamma_T \downarrow$	$\eta_{C/T} \downarrow$	~1:2.0
Blade rubbing	$\Gamma_C \downarrow$ & $\Gamma_T \uparrow$	$\eta_{C/T} \downarrow$	~1:2.0

Gas Turbine Degradation Simulation

Before starting any degradation simulations it is necessary to establish the base line (design point performance) of the plant. This base line performance point is represented by (0.0) value on all deterioration graphs shown below. Once the design point has been identified, then the magnitude of faults that represents a physical fault of the component in consideration, mentioned in previous section, to be implanted on each component has to be established.

Unfortunately, although there is a lot of work published on the subject of gas turbine performance deterioration, [1–4], the applied degradation magnitude to each component, when simulating gas turbines deterioration performance, in most cases is either arbitrary or based on some published experimental results. Therefore, in present study the values mentioned by [1] and [5] were taken as a guidelines from which the implanted faults were estimated. Table 2, [5], with some modifications, shows a summary of how component isentropic efficiency changes vary with degradation. These values were applied in all calculations to the appropriate components.

Throughout this work it was assumed that there was no equipment washing or any type of maintenance taken on the gas turbine until the deterioration reaches 5% from the original design point performance.

Gas Turbine Degradation Simulation Results

The most important gas turbine deterioration simulation results are briefly presented graphically in Figs. 2 through 5. It is worth noting here that the term (compressor degradation) used in all figures means either compressor fouling or erosion, or both.

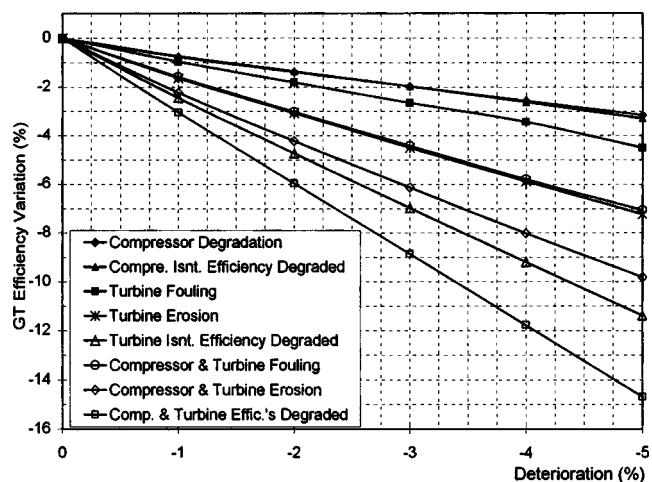


Fig. 2 GT efficiency variation with component deterioration

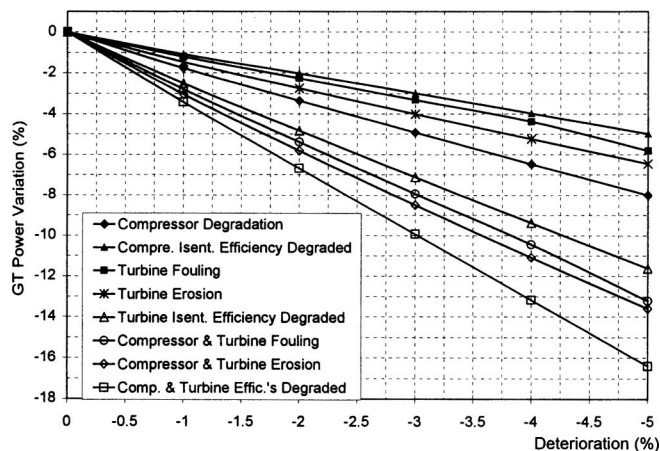


Fig. 3 GT power variation with component deterioration

Figure 2 shows the degradation effects on gas turbine efficiency. It can be observed that the compressor degradation whether it be fouling and/or erosion gave the lowest deterioration in gas turbine overall efficiency. The value of efficiency deterioration was about -3.1% with compressor degradation by 5% . This implies that, gas turbine overall efficiency is less sensitive to compressor degradation than to the degradation of other components, as it will be shown later.

On the other hand, as expected, combined compressor and turbine isentropic efficiencies degradation gave the most severe effect in overall gas turbine efficiency. As it can be seen, the plant efficiency deterioration was approximately -14.6% (from original design point) with only 5% degradation.

Although the expectation was to see more reduction in plant thermal efficiency when both compressor and turbine are either fouled or eroded together as this fault implies a reduction in both component isentropic efficiency in addition to fouling or erosion, the results obtained was not so. This can be explained as follows,

1. It appears that the ratio of isentropic efficiency degradation to the component fouling or erosion applied in each case when simulating component fouling or erosion is $0.5:1.0$ (see Table 1). This means that the value of component isentropic efficiency degradation along with 5% of fouling or erosion is 2.5% . Whereas in

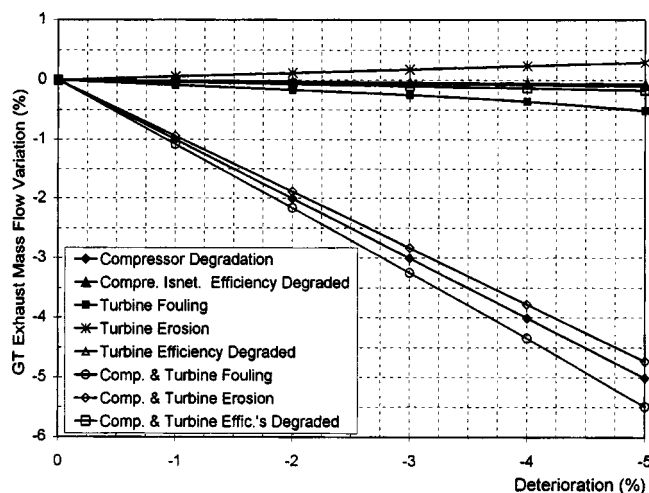


Fig. 4 GT exhaust mass flow variation with component deterioration

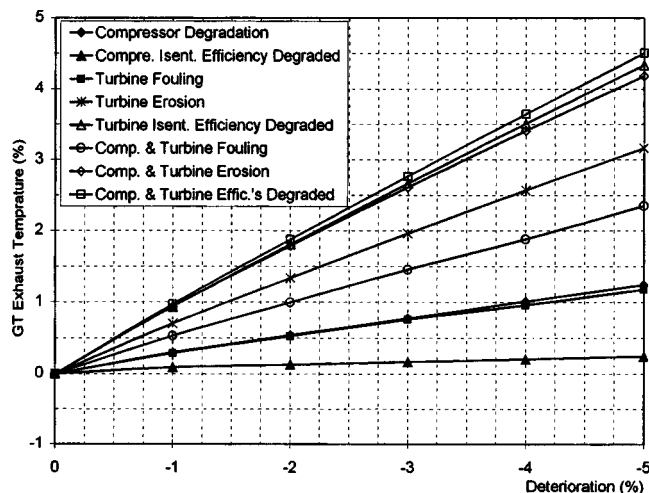


Fig. 5 GT exhaust temperature variation with component deterioration

the situation of simulating the same case (5% degradation) of only component isentropic efficiency degradation the value used was 5% .

2. Secondly, In the case of both components fouling, the increased pressure ratio through the turbine caused by fouling increased the power output of the turbine. This led to a higher plant power output and a higher thermal efficiency than the case where only the component isentropic efficiency was degraded.

3. Similarly the combination of decreased compressor mass flow with an increased turbine flow capacity, due to erosion by 5% , led to a higher (about 2.8%) reduction in the plant's overall efficiency in comparison to the case mentioned above (fouling in both components). This is exactly the opposite of what was mentioned earlier. I.e., the decreased pressure ratio through the turbine resulted in a lower power output of the turbine, and hence a reduced overall power output of the engine which then reflected on the engine's overall efficiency.

4. Finally, this also applies in the case where only the component isentropic efficiencies are reduced, the mass flow capacity is approximately at its DP value and hence the same pressure ratio approximately. The obvious result of this shortfall in component efficiencies will be a reduction in the overall power output and efficiency of the plant.

Figure 3 below shows different component degradation effects have on gas turbine power output. The compressor inlet mass flow is reduced by 5% due to degradation. This resulted in a reduction of about 8% in engine power output. This is similar to that observed elsewhere, [1], which was 7% . It is worth mentioning here that, in this case of Diakunchak [1] the compressor isentropic efficiency at fouling of 5% was 1.8% , while in the present case it was assumed to be 2.5% . This led to a reduction in engine power output, as stated above, by 1% higher than the one quoted, [1].

Figures 2 and 3 show that the degradation in turbine isentropic efficiency has a higher effect on engine efficiency and power than that of compressor isentropic efficiency. A compressor isentropic efficiency degradation of 5% led to a reduction of about 4.9% in engine power with a corresponding reduction of about 3.3% in engine efficiency. The same amount of degradation in turbine isentropic efficiency (5%) resulted in about 11.6% fall in engine power, and about 11.3% reduction in engine efficiency.

This is due to the fact that, the turbine has two tasks, one to drive the compressor, and the second to produce the engine net power output, and hence the turbine degradation reflection is higher because of the much larger power involved.

A close look at Figs. 2 and 3 will show that, the degradation of 1% in turbine isentropic efficiency has led to approximately 2.5%

reduction in both gas turbine power output and efficiency. These values are in agreement with those described elsewhere, [1].

In the case of a simple Brayton cycle operating in isolation the change in engine exhaust gas properties due to component degradation could be neglected as the gases are usually discharged to the stack. However, in the case of combined cycle, where the bottoming (steam) cycle is dependent on the exhaust gases of the topping (Brayton) cycle, it is very important to know the exhaust gas properties and how they change with different components degradation because there is a strong relationship between gas turbine exhaust properties and steam turbine output in a combined cycle.

The effects of different gas turbine component degradation on engine exhaust mass flow and temperature are illustrated in Figs. 4 and 5, respectively. As Fig. 4 shows, it can be seen clearly that the exhaust mass flow has increased slightly (about 0.3%) with 5% of turbine erosion. On the other hand, in the case of compressor and turbine erosion combination, the process is solely a function of compressor erosion, as the turbine mass flow would follow the compressor inlet mass flow. However, as this figure shows, the exhaust mass flow due to compressor and turbine erosion combination by 5% is higher than that of the compressor degradation (erosion) alone by about 0.28%. This is because of the increased mass flow at the turbine inlet due to erosion. It is also clear from this figure that the compressor and turbine fouling combination has the highest severe effect on exhaust mass flow. The effect of any of the components isentropic efficiencies degradation on gas turbine exhaust mass flow, as can be seen from this figure as well, is approximately negligible.

Figure 5 below shows the variation of gas turbine engine exhaust temperature with different components degradation. Contrary to the effects of components isentropic efficiencies degradation on gas turbine exhaust mass flow, it can be noted that the combination of compressor and turbine isentropic efficiency degradation has the highest effect on exhaust temperature.

As this figure shows, the increase in exhaust temperature was about 4.5%, from original DP, with 5% degradation in both compressor and turbine isentropic efficiencies (notice the small share of the compressor isentropic efficiency degradation). This is due to the fact that for a constant TET, the reduced turbine isentropic efficiency would mean a lower power output, and hence a higher amount of heat content rejected at the exhaust gases, and thus a higher exhaust temperature. This could be favorable to the steam turbine plant, as it would lead to an increase in the steam production, and hence increased steam turbine power output. However, the decreased gas turbine power due to this fault may lead, depending on steam turbine power output production, to an outcome reduction in CCGT power output.

Steam Turbine Performance Deterioration

Given that the focus of the analysis described in this paper is the effect of gas turbine component degradation on the combined cycle, all steam turbine components were assumed to be operating in their clean or new condition, regardless of the situation of the gas turbine plant. The effects of the deteriorated topping (gas turbine) cycle on the steam cycle is discussed here. A more detailed discussion of the effects of degraded components of steam turbine plant have on its and on CCGT power plant performance, as the bottoming cycle of CCGT plant, can be found in [6].

As Fig. 1 shows, the steam section of the plant consists of a single pressure HRSG, one steam turbine, generator, condenser, and a feed pump. This plant has the following specifications at design point conditions:

live steam pressure	= 65.4 bar
live steam temperature	= 537.8°C
steam mass flow	= 67 kg/sec
steam turbine isentropic efficiency	= 89.48%
superheater surface area	= 8424.8 m ²
evaporator surface area	= 29315.6 m ²
economiser surface area	= 38004.1 m ²
condenser surface area	= 3942.9 m ²
HRSG efficiency	= 81.11%
steam turbine plant power output	= 76454.14 kW
steam turbine plant efficiency	= 33.97%

Steam Cycle Performance Simulation

First, the design point performance of the steam turbine plant must be established. Here the components of the steam (bottoming) plant were kept in their original, clean, condition. The exhaust gases, at different deterioration conditions of gas turbine, received from the upper cycle are then passed through bottoming cycle, and the results were observed. By doing so, it was possible to investigate how the bottoming cycle behaves with different types of gas turbine degradations. This is done (as already mentioned) by utilizing a Fortran code produced, at Cranfield University, especially for this purpose (see [6]).

Steam Turbine Operating Performance

As with gas turbine simulation, the most important steam turbine deterioration simulation results are represented graphically in Figs. 6 through 8. It is worth reminding the reader here that the point (0.0) on (Y) axes of all graphs means the design point value.

Figure 6 shows the variation of steam production in the HRSG with different types of gas turbine component degradations. It is clear that all types of gas turbine degradation have led to an increase in the steam mass flow, except compressor degradation.

Although, as Fig. 5 shows, there was an increase in exhaust temperature due to compressor degradation which is expected to result in an increase in steam production in the HRSG, the decrease in gas turbine exhaust mass flow caused by the same fault, as Fig. 4 shows, was predominant, and hence the outcome was a decrease in steam mass flow. Another observation from this figure is that the steam production in the HRSG was more sensitive to the faults caused by GT turbine, rather than to those caused by the compressor. This is because (as this figure shows) GT turbine erosion and isentropic efficiency degradations, as individual faults, resulted in the highest amount of steam production. By comparing this with Figs. 4 and 5, it will be observed that the

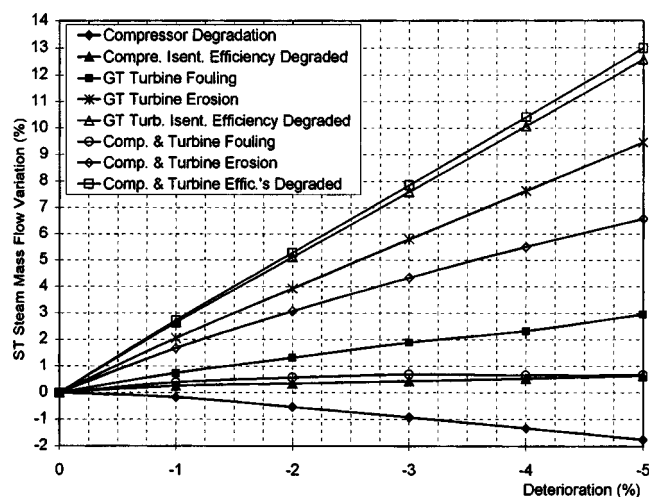


Fig. 6 ST steam mass flow variation with gas turbine component deterioration

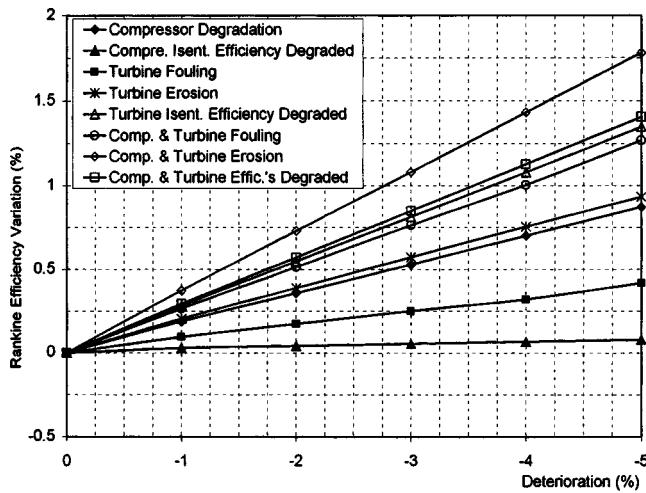


Fig. 7 Rankine efficiency variation with gas turbine component deterioration

exhaust temperature (Fig. 5) was the main contributor to this increase of steam mass flow production, as exhaust mass flow was almost constant with GT turbine erosion and isentropic efficiency degradation (see Fig. 4).

Finally, the increase in steam mass production due to compressor and turbine isentropic efficiencies degradation of 5% together was about 13.0 percent. As Fig. 6 shows, this value is approximately 0.5% above that one caused by the turbine isentropic efficiency degradation. Therefore, this 0.5% increase is mainly the contribution of compressor isentropic efficiency degradation.

The next important performance parameter to discuss here is the steam turbine plant (Rankine) efficiency variation with gas turbine degradation, which is illustrated in Fig. 7 below. The thermal efficiency definition of steam turbine (bottoming) plant is given by

$$\eta_R = \frac{W_{SC}}{Q_{HRSG}} \quad (2)$$

This equation shows that, the steam turbine cycle efficiency is a function of steam turbine net power output and the heat transferred in the HRSG. Now by looking at Fig. 9 it will be seen that all types of GT degradations resulted into an increase in the

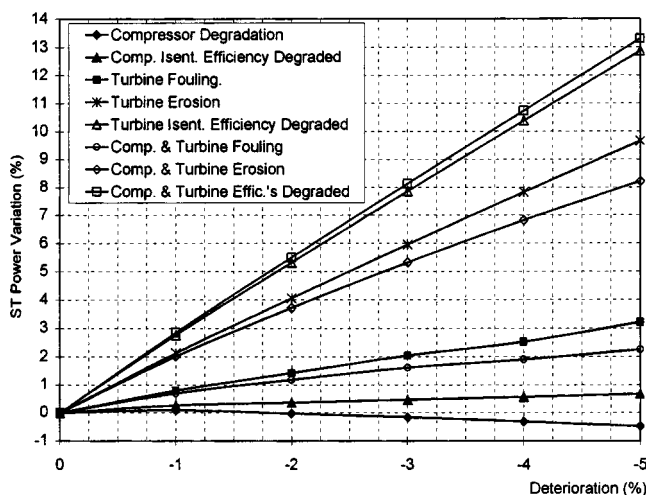


Fig. 8 ST power variation with gas turbine component deterioration

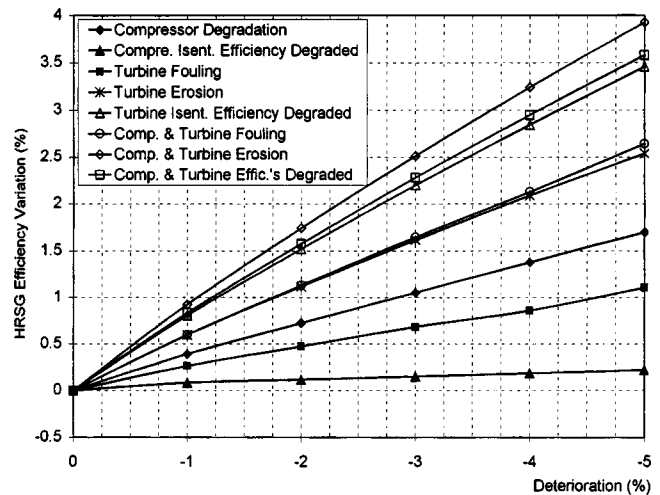


Fig. 9 HRSG efficiency variation with gas turbine component deterioration

HRSG efficiency. This clearly explains the behavior trend of the graphs representing the steam plant efficiency that appeared on Fig. 7.

Another useful observation is made by comparing Figs. 4 and 5 with Fig. 9. The GT exhaust temperature has a larger effect on steam cycle efficiency, (which by itself is a function of HRSG efficiency), over the exhaust mass flow.

Steam turbine power is a function of steam mass flow and its enthalpy. By comparing Fig. 8 with Fig. 6 it can be noticed that the steam turbine power is more or less following the mass flow behavior. The steam production in the HRSG is more sensitive to the faults of the gas turbine turbine, rather than those of the compressor. Therefore, since steam turbine power is a function of steam mass flow, the steam turbine power is more sensitive to the faults in the GT turbine, rather than to those in the compressor.

Figure 8 shows how steam turbine power varies with different types of gas turbine degradation. The higher exhaust temperature caused by degradation increased the steam turbine power output. As shown in Fig. 8, the maximum effect of gas turbine degradation on steam turbine power output was in the case of compressor and turbine efficiencies degradation. The increase in steam turbine power output was as high as 13.3% with compressor and turbine efficiencies degraded by 5%.

Combined Cycle Degradation Results

A FORTRAN code was specially developed to carry out the degraded performance analysis of the bottoming cycle at design and off design conditions. Since current study deals with a single pressure CCGT plant, only heat transfer paths shown by Fig. 10 (in addition to steam turbine section) were dealt with. The most important performance deterioration simulation results of combined cycle are shown in Figs. 11, 12, and 13. As Eq. (3) below shows, the combined cycle efficiency is a function of gas turbine cycle efficiency, HRSG efficiency, and steam cycle efficiency.

$$\eta_{CC} = \eta_{GT} + (1 - \eta_{GT}) \cdot \eta_{HRSG} \cdot \eta_{SC} \quad (3)$$

Even though, as Figs. 7 and 9 shows, there was an increase in steam cycle and HRSG efficiencies with some (same) cases of GT component degradation (e.g., GT turbine erosion), still the effect of decreased GT efficiency (see Fig. 2) has a higher effect on CCGT efficiency. This is shown in Fig. 11. The combined cycle efficiency has fallen with all types of GT component degradations, even with steam turbine plant components at their original (DP) conditions. The GT efficiency has a predominant effect on CCGT power plant efficiency over steam cycle and HRSG efficiencies.

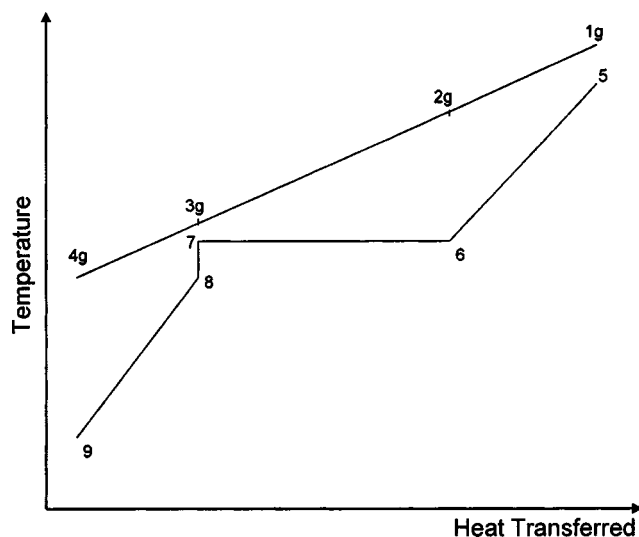


Fig. 10 Heat transfer diagram for a single pressure CCGT power plant

Figure 11 also shows that the CCGT efficiency is more sensitive to the faults that caused by GT turbine, rather than those caused by the compressor. The GT turbine isentropic efficiency degradation (as an individual fault) gave the highest value in CCGT efficiency reduction. The efficiency deterioration was about 3.5% with 5% degradation in GT turbine isentropic efficiency. When both isentropic efficiencies of GT turbine and compressor are degraded together by 5%, the reduction in CCGT efficiency was about 5.0%.

Given that, typically, the gas turbine output is approximately two thirds of the total output of an unfired CCGT, it is expected to see a stronger influence on CCGT performance arising from changes in the gas turbine than in the steam cycle. Comparing Fig. 12 with Figs. 3 and 8 it may be noticed that the CCGT power output is more or less following the behavior of the gas turbine's power output. This shows, in addition to the conclusion withdrawn on CCGT efficiency above, that the behavior of CCGT plant performance is affected more by gas turbine conditions than by steam turbine conditions. As this figure also shows, although only gas turbine performance was deteriorated while steam turbine was kept at its original DP conditions, the net outcome was a reduction in CCGT plant power output. This reduction was at its

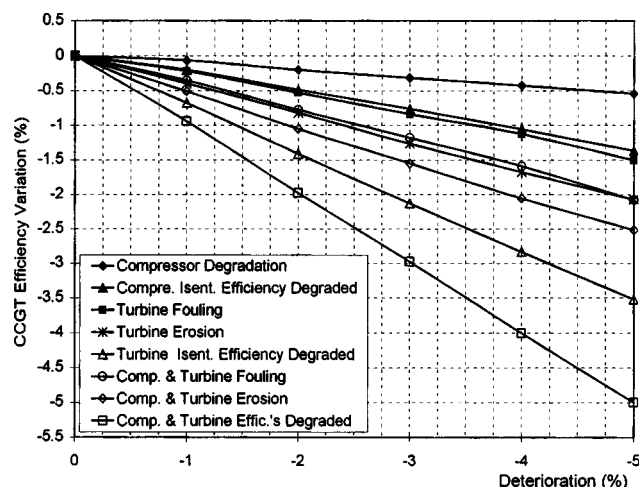


Fig. 11 CCGT efficiency variation with gas turbine component deterioration

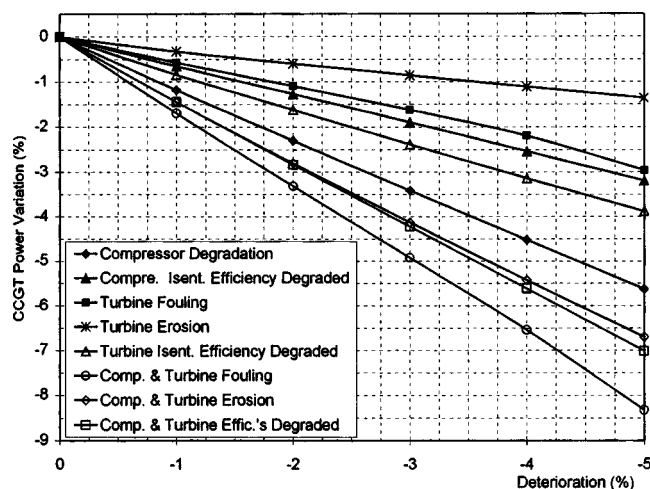


Fig. 12 CCGT power variation with gas turbine component deterioration

highest with the combination of compressor and GT turbine fouling. This was approximately 8.3% with both components fouling together by 5%.

Another important CCGT performance parameter is the stack temperature. This is mainly a measure of the amount of gas turbine exhaust heat utilisation by the bottoming cycle. By definition, HRSG efficiency is a function of stack temperature and exhaust inlet temperature for a given ambient temperature, Eq. (4).

$$\eta_{\text{HRSG}} = \frac{T_{1g} - T_{4g}}{T_{1g} - T_{\text{amb}}} \quad (4^1)$$

This equation shows that, for a given inlet exhaust and ambient temperatures, the HRSG efficiency increases with decreased stack temperature (T_{4g}) and vice versa. This can be clearly perceived by comparing Figs. 9 and 13. Now by definition as Eq. (3) shows, CCGT efficiency is a function of HRSG efficiency. Therefore, as Fig. 13 shows, the stack temperature was at its lowest value (about -8.9%) with compressor and turbine erosion by 5%. This reduction in stack temperature, as Fig. 9 shows, was reflected on the HRSG efficiency.

¹See Fig. 10 for the definition of the notations used in this equation.

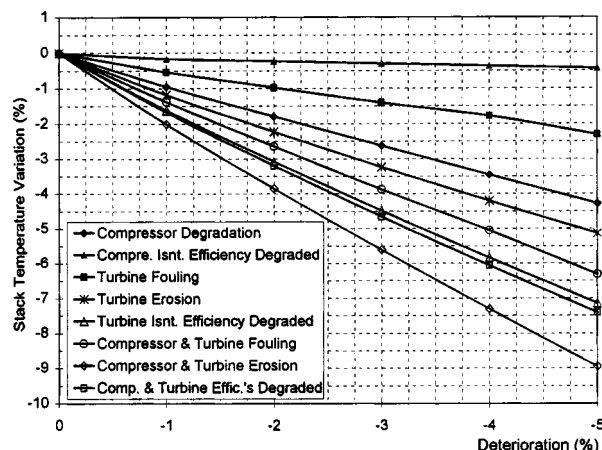


Fig. 13 Stack temperature variation with gas turbine component deterioration

Conclusions

The results obtained from gas turbine simulation Software at Cranfield University were very similar, in fact in many cases are exact, to those quoted elsewhere, [1].

As an individual fault, the GT turbine isentropic efficiency degradation has the most severe effect on gas turbine power and efficiency.

Gas turbine exhaust mass flow is merely a function of flow capacity through the engine; i.e., decreases with fouling, and increases with erosion. In addition, the effect of any of the components isentropic efficiencies degradation on gas turbine exhaust mass flow is almost negligible.

Steam turbine cycle steam mass flow, and hence steam turbine power output are more sensitive to the faults that caused by GT turbine, rather than to those caused by GT compressor.

The GT exhaust temperature has a predominant effect on steam cycle efficiency over the GT exhaust mass flow.

Among the three interrelated CCGT efficiency parameters that shown in Eq. (3), the gas turbine turbine efficiency was the most important parameter on CCGT efficiency.

The behavior of CCGT plant performance is more affected by gas turbine cycle conditions than by steam turbine cycle conditions.

Nomenclature

A	=	area
CCGT	=	combined cycle gas turbine plant
DP	=	design point
FOD	=	foreign object damage
GT	=	gas turbine
HRSG	=	heat recovery steam generator
OD	=	off-design condition
P	=	Pressure
Q	=	Heat Transfer

T	=	Temperature
TET	=	turbine entry temperature
$\dot{w}$	=	Inlet mass flow
W	=	power
Γ	=	nondimensional mass flow
η	=	efficiency

Subscripts

amb	=	ambient
c	=	compressor
CC	=	combined Cycle
C/T	=	compressor and/or turbine
g	=	gas
HRSG	=	heat recovery steam generator
i	=	inlet
R	=	Rankine cycle
SC	=	steam cycle
T	=	turbine

References

- [1] Diakunchak, I. S., 1992, "Performance Deterioration in Industrial Gas Turbines," ASME J. Eng. Gas Turbines Power, **114**(2).
- [2] Lakshminarasimha, A. N., Boyce, M. P., and Meher-Homji, C. B., 1994, "Modelling and Analysis of Gas Turbine Performance Deterioration," ASME J. Eng. Gas Turbines Power, **116**(1).
- [3] Tabakoff, W., 1986, "Compressor Erosion and Performance Deterioration," AIAA/ASME 4th Joint Fluid Mechanics, Plasma Dynamics, and Laser Conference, Atlanta, GA, May 12–14.
- [4] Tabakoff, W., Lakshminarasimha, A. N., and Pasin, M., 1990, "Simulation of Compressor Performance Deterioration Due to Erosion," ASME J. Eng. Gas Turbines Power, **112**(1).
- [5] Escher, P. C., 1995, "Pythia: An Object-Oriented Gas Path Analysis Computer Program for General Applications," Ph.D. thesis, Cranfield University, School of Mechanical Engineering, UK.
- [6] Zwebek, A. I., and Pilidis, P., 2001, "Degradation Effects on Combined Cycle Power Plant Performance, Part II: Steam Turbine Cycle Component Degradation Effects," ASME Paper No. 2001-GT-389.

Degradation Effects on Combined Cycle Power Plant Performance— Part II: Steam Turbine Cycle Component Degradation Effects

A. Zwebek

e-mail: gotross@hotmail.com

P. Pilidis

e-mail: p.pilidis@cranfield.ac.uk

Department of Power Engineering and
Propulsion,
School of Engineering,
Cranfield University,
Bedford MK43 0AL, UK

This is the second paper exploring the effects of the degradation of different components on combined cycle gas turbine (CCGT) plant performance. This paper investigates the effects of degraded steam path components of steam turbine (bottoming) cycle have on CCGT power plant performance. Areas looked at were, steam turbine fouling, steam turbine erosion, heat recovery steam generator degradation (scaling and/or ashes deposition), and condenser degradation. The effect of gas turbine back-pressure on plant performance due to HRSG degradation is also discussed. A general simulation FORTRAN code was developed for the purpose of this study. This program can calculate the CCGT plant design point performance, off-design plant performance, and plant deterioration performance. The results obtained are presented in a graphical form and discussed.

[DOI: 10.1115/1.1519272]

Introduction

It is the nature of power plants that they do not work at all times at their design point conditions. From the experience built up over the past years of operation of such plants it was observed that two main sources are available which offsets the plant from its design point conditions, these are

1. off-design due to normal conditions (changes of ambient conditions and part load) and
2. off-design due to abnormal conditions (changes in fluid path components configuration due to degradation).

While the first problem can be dealt with up to some extent, the second one is really hard to control and it is a function of many interrelated parameters.

According to [1] and others, fouling of heat exchanger surfaces in power plants results in huge economical losses. Von Nostrand [2] estimated the total cost of fouling of heat transfer surfaces for petroleum refining in the non-Communist countries as high as \$4.41 billion per year.

Although it has been recognized since, a long time ago the effect of degradation of heat exchanges have on heat transfer, there seems little field data or experimental research work found on this subject in the open literature, [1] and [3].

It is well known that the efficiency of the steam turbine (bottoming) cycle as a part of CCGT plant is primarily a function of GT efficiency, [4], and HRSG efficiency. Therefore, it is very important to see how degraded HRSG affects the steam turbine power plant performance, and hence CCGT plant performance. For the effects of gas turbine component degradation see [4].

Therefore it is becoming of great importance to predict in advance the behavior of thermal plants as they get older and older so as plans can then be made in advance to avoid long shutdowns and hence a lot of economical losses. Also this prediction of plant behavior helps in pre-arranging maintenance plans.

This paper describes how common faults affect CCGT plant performance. In this study, it is assumed that the gas turbine is

working at its original design point conditions, regardless of its lifetime. Then by applying a certain amount, as described below, of degradation to different components of the steam (bottoming) cycle the performance behavior of the plant accordingly could have been simulated. In this way the plant's behavior can be monitored.

This study is performed by way of simulation. The gas turbine degradation effects on a CCGT plant were investigated by using Turbomatch (a FORTRAN code available at Cranfield University), [4]. To simulate the degradation effects of steam turbine (bottoming) cycle on the CCGT plant; a new FORTRAN code was developed. The obtained results are then discussed and compared with published data wherever possible.

Gas Turbine Performance Deterioration

Contrary to the first paper, [4], the strategy throughout this paper is to keep the gas turbine at its design point conditions while applying different amounts of degradation to the steam turbine cycle components. The only one exception case to this is the simulation of increased back-pressure of the gas turbine as a result of HRSG heat transfer surfaces fouling (see the discussion below).

To have comparable simulation results of CCGT plant components degradation as a whole, it was meant to use the same gas turbine (topping) and steam turbine (bottoming) cycle plants (Fig. 1) that were used in the previous paper, [4]. This helps in having a global overview of the plant's behavior in accordance with different component degradation. It also helps in investigating the effects of degradation of both plants as separate units. And hence identify the sensitive parts of both plants to degradation. Also by comparing the results obtained in both cases a wider understanding of the response of CCGT plants to different component degradation can be observed.

Therefore, main gas path components of the gas turbine, namely compressor, combustion chamber, and turbine, were assumed that they are working at their design point conditions at all times.

In this study a typical gas turbine having the following design point conditions, [4], was used:

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, New Orleans, LA, June 4–7, 2001; Paper 2001-GT-0389. Manuscript received by IGTI, December 2000, final revision, March 2001. Associate Editor: R. Natole.

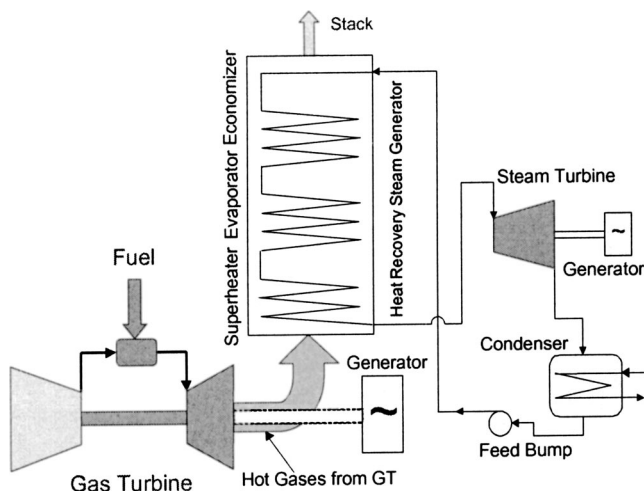


Fig. 1 Schematic diagram of a single pressure CCGT power plant (Zwebek and Pilidis [4])

inlet mass flow	= 408.66 kg/sec
compressor pressure ratio	= 15.2
turbine entry temperature	= 1697.80 K
exhaust mass flow	= 419.4 kg/sec
exhaust temperature	= 871.24 K
net power output	= 165.93 MW
thermal efficiency	= 35.57%

Gas Turbine Degradation Simulation

Before starting any degradation simulations it is necessary to establish the base line (design point) performance of the plant. This base line performance, as in [4], is represented by (0.0) value on all deterioration graphs shown below. Once the design point has been identified, then the magnitude of faults that represents a physical fault of the component in consideration to be implanted on each component has to be established.

Unfortunately, not much literature was found on the subject of CCGT plant degradation, or on modeling of this problem, including the effect of back-pressure. Therefore to simulate the effect of back-pressure on gas turbine performance, due to HRSG degradation some assumptions has been made. According to [5], an increase in back-pressure by ≈ 0.0025 atm results in a reduction in gas turbine power by $\approx 0.3\%$. Because of the inherent problems which accompanies the increase of back-pressure, e.g., high torque on the shaft, coupling forces on thrust bearing, and vibration, it was assumed that maximum it can go up to 0.025 atm over the DP value. Reference [5] stated that typical back-pressure ranges from 0.025 to 0.037 atm above the design value.

Gas Turbine Degradation Simulation Results

As already mentioned above, in this paper the gas turbine of the current CCGT plant was kept at its DP conditions, but because of the increased back-pressure due to HRSG degradation, the gas turbine will not work at its DP condition any more.

As Fig. 2 shows, a back-pressure increase of only 2.5% resulted in a reduction of gas turbine thermal efficiency and power of approximately 1.7%, while the exhaust mass flow was almost constant and the exhaust temperature increased by about 0.65%.

Steam Cycle Performance Deterioration

As with the case of gas turbine gas path components, [4], the steam turbine cycle steam path components are also subjected to degradation due to fouling, erosion, and/or corrosion.

While in the case of the gas turbine almost only one surface (outer) is subjected to these degradation effects, which simplifies,

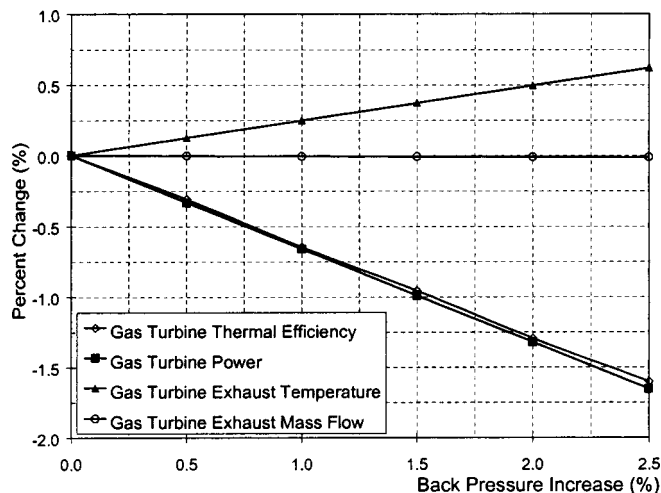


Fig. 2 Back pressure effects on GT performance

up to some extent, the process of investigating this problem. In the case of a steam turbine plant the degradation is affecting two surfaces of some components of the plant at the same time. In the case of the HRSG and the condenser (heat exchangers) two types of degradation are available, one is the outer tubes surface fouling and corrosion, and another is the innertubes surface scaling or erosion. This is, of course, in addition to the degradation of steam turbine unit.

Even with the latest fuel treatment techniques, the exhaust gases from the gas turbine will contain some chemicals in a form of flying ash and soot which deposits on the outer heat transfer surfaces of the HRSG. Also, the impurities, however, water treatment techniques are used in circulating water deposits on the inner walls of the heat exchanger pipes. These then lead to a reduction in the heat exchanger performance (effectiveness). The condenser will also behave in a similar way. The third cause that leads to steam cycle plant performance deterioration comes from the steam turbine unit degradation.

In order to cover most types of degradation that might attack the bottoming cycle of the CCGT plant, it was assumed that each component might degrade separately. Then all components were assumed to degrade together. This helps to establish the nature of the faults and to assess if they are additive or not. The faults investigated were the following:

1. economizer degradation,
2. evaporator degradation,
3. superheater degradation,
4. steam turbine fouling,
5. steam turbine erosion,
6. ST isentropic efficiency degradation,
7. condenser degradation,
8. combination of all faults mentioned above, and
9. gas turbine back-pressure increase due to heat exchanger (HRSG) surfaces fouling.

Fault Representation

In order to investigate the effects of faults mentioned in the previous section on the steam turbine plant performance as a stand alone unit, and hence on the CCGT plant as a whole, these faults were fed into the program as a percent reduction of the original design point value. This is done as follows:

(i) **Heat Exchanger Degradation.** The degradation of either of the heat exchangers (economizer, evaporator, superheater, and condenser) was simulated by assuming a percent reduction in the original DP value of the overall heat transfer coefficient of the heat exchanger in concern.

Table 1 Representation of component degradation

Fault	Represented by	Range
Gas turbine back-pressure	GT Back-pressure rise	0.0–(–2.5%)
Economizer degradation	Drop in U	0.0–(–5.0%)
Evaporator degradation	Drop in U	0.0–(–5.0%)
Superheater degradation	Drop in U	0.0–(–5.0%)
Condenser degradation	Drop in U	0.0–(–5.0%)
Steam turbine fouling	Drop in Γ	0.0–(–5.0%)
	Drop in η_T	0.0–(–2.5%)
Steam turbine erosion	Rise in Γ	0.0–(+5.0%)
	Drop in η_T	0.0–(–2.5%)
FOD	Drop in η_T	0.0–(–5.0%)

(ii) **Steam Turbine Fouling.** The steam turbine fouling is represented by a reduced flow capacity at the inlet of the turbine plus a reduction in turbine isentropic efficiency. By doing so, it is assumed that there is a blockage in the inlet area of the turbine due to particles deposition accompanied by a decrease in its isentropic efficiency due to surface roughness, for example.

(iii) **Steam Turbine Erosion.** The steam turbine erosion is represented by increasing mass flow capacity at the turbine inlet and at the same a reduction in turbine isentropic efficiency.

The two previously mentioned phenomenon (fouling and erosion) are represented by changing the so-called nondimensional mass flow (Eq. (1)) of the component map.

$$\frac{\dot{W}\sqrt{T_i}}{PA} = \text{constant}. \quad (1)$$

(iv) **Component Efficiency Degradation.** This fault affects the steam turbine unit. It is modeled by reducing the unit's isentropic efficiency of the appropriate map and keeping all other map parameters at their normal condition. In this case, it was assumed that the component isentropic efficiency may decrease from its DP value due to any reason, such as blade tip rubs.

(v) **Gas Turbine Back Pressure.** The increased back-pressure at the gas turbine exhaust is represented as an increase in the GT exhaust outlet pressure.

The above-mentioned faults are applied to different components of the plant in different values. Table 1 summarizes these faults and their ranges at which they were applied to each component.

Steam Cycle Degradation Simulation

Before starting any degradation simulation it was necessary to establish a datum working line of the plant. Therefore, by using the developed code, a steam cycle DP was arrived at. This cycle was having the following DP conditions, [4]:

live steam pressure	= 65.4 bar
live steam temperature	= 537.8°C
steam mass flow	= 67 kg/sec
steam turbine isentropic efficiency	= 89.48%
superheater surface area	= 8424.8 m ²
evaporator surface area	= 29315.6 m ²
economizer surface area	= 38004.1 m ²
condenser surface area	= 3942.9 m ²
HRSg efficiency	= 81.11%
steam turbine plant power output	= 76454.1 kW
steam turbine plant efficiency	= 33.97%

Having established these DP conditions, the steam cycle was then analyzed in a degraded mode. The amount of degradation applied to each component was really a matter of assumption as there was no such documented work of the similar type in the open literature. Therefore, in the present work, when modeling steam turbine unit, it was assumed that every 1.0% deterioration in mass flow capacity (fouling or erosion) would result in a deterioration of (0.50%) in steam turbine isentropic efficiency. In the

case of HRSg degradation simulation, it was assumed that every 1.0% deterioration in all components (economizer, evaporator, and superheater) would result in an increase in the gas turbine back-pressure by 0.5% (see Fig. 2).

Throughout this work it was assumed that there was no equipment washing or any type of maintenance taken on the steam turbine cycle plant until the deterioration reached 5% from the original design point performance.

Steam Cycle Degradation Simulation Results

The most important steam turbine cycle deterioration simulation results are represented graphically in Figs. 3 through 7. It is worth reminding the reader here that the point (0.0) on (Y) axes of all graphs represents the design point value.

Figure 3 below shows the steam turbine power variation with different components degradation. As it can be seen from this figure, the steam turbine unit isentropic efficiency (as an individual fault) was the supreme affecting parameter on the plant's power. When the steam turbine unit isentropic efficiency degraded by 5.0% the reduction in ST power output was about 5.6%. The power deterioration due to steam turbine fouling or erosion, as this figure shows, is less in value than in the case of steam turbine unit isentropic efficiency degradation alone. The reason for this is that the ratio of isentropic efficiency degradation to the ST fouling and/or erosion applied in each case when simulating ST fouling or erosion is 0.5:1.0. This means that the value of ST isentropic efficiency degradation along with 5% of fouling or erosion is 2.5%. Whereas in the case of simulating the ST isentropic efficiency degradation individually with 5.0% degradation, the value used was 5.0%.

It is well known that the ST power is a function of steam mass flow. Now by looking at Fig. 5 it will be observed that evaporator degradation, compared to economizer and superheater degradations, resulted in the highest level of steam mass flow reduction, which then resulted in reducing the plant's power output.

Also as this figure shows, the erosion effects on ST cycle power output was predominant over the effect of steam turbine fouling. In the case of steam turbine fouling by 5.0% the deterioration in ST power was about 2.6%, where as in the case of steam turbine erosion by the same amount (5.0%) the deterioration in ST power was in the region of 3.3%. The reason for this is that in the case of steam turbine fouling the increase in the inlet pressure due to inlet area blockage did compensated for some of the power loss and hence resulted in a lower reduction in ST power output compared to ST erosion.

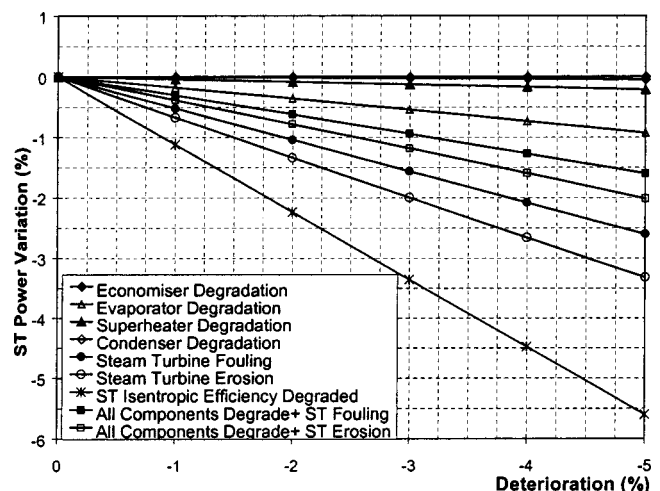


Fig. 3 Steam turbine power variation with component deterioration

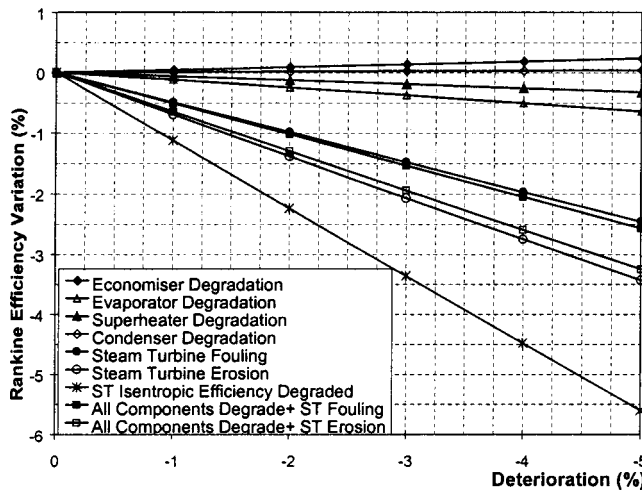


Fig. 4 Rankine efficiency variation with component deterioration

The GT efficiency has a predominant effect on CCGT power plant efficiency over steam cycle and HRSG efficiencies.

Figure 4 shows the variation of steam cycle (Rankine) efficiency with different types of steam cycle component degradations. In general by comparing Figs. 3 and 4 it will be noticed that the ST cycle thermal efficiency is more or less following the plant's power, since efficiency is a function of power output.

As this figure also shows, while Rankine efficiency is almost constant with condenser and economizer degradations, it showed a little decrease (about -0.32%) with degraded superheater by 5.0%. The largest deterioration in Rankine efficiency due to degradation of any of the three components of the HRSG was encountered with evaporator degradation. This was about 0.64% efficiency deterioration with 5.0% degradation.

As an individual fault, the steam turbine fouling by 5.0% gave about 2.5% reduction in steam cycle efficiency. When added to this the degradation of other components, the increase in Rankine efficiency deterioration was about 0.11%. A similar result was obtained in the case of steam turbine erosion simulation. Among all the faults investigated, the steam turbine isentropic efficiency degradation gave the highest level in Rankine efficiency deterioration. Sanders [6] stated that the degradation of state line efficiency of 1.0% in each section of the steam turbine unit would

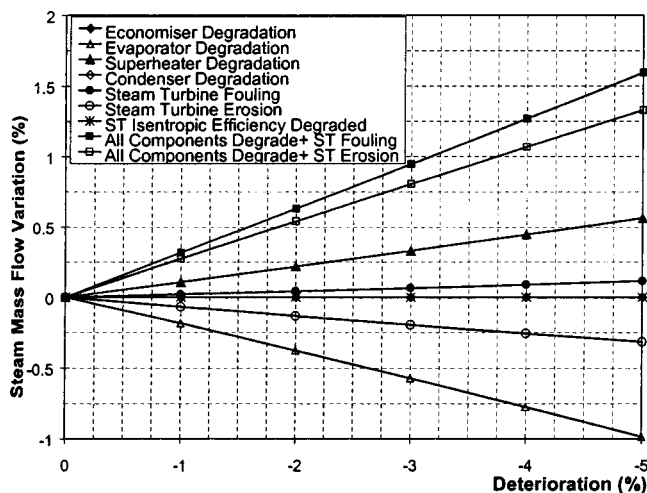


Fig. 5 ST steam mass flow variation with steam turbine component deterioration

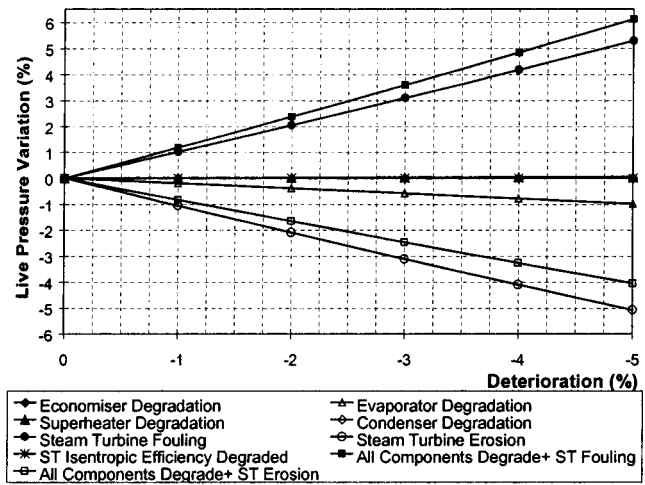


Fig. 6 Live steam pressure variation with steam turbine component deterioration

result in approximately 1.0% deterioration in cycle heat rate. Now, as it can be seen from Fig. 4, the steam cycle efficiency thermal deterioration was approximately 5.6% with 5.0% degradation of steam turbine unit isentropic efficiency degradation. This leads us to two conclusions:

1. The steam turbine isentropic efficiency as a performance parameter has the highest effect on Rankine efficiency.

2. The simulation results obtained from the developed code are in agreement with what the authors put their hands on in the open literature. This gives a sensible confident in the code developed at Cranfield University, which is the source of simulation result of this paper.

By looking at both Figs. 3 and 4 at the same time, it will be observed that the evaporator degradation was the most affecting fault on ST cycle deterioration. This is due to the fact that the evaporator is producing the largest duty, and hence the reflection of its degradation on the cycle was the highest.

Figures 6 and 7 shows how live steam pressure and temperature varies with different component degradation. As Fig. 6 shows the blockage of the steam turbine inlet due to fouling by 5.0% resulted in about 5.3% increase in live steam pressure at the ST turbine inlet. The combination of all other types of degradation

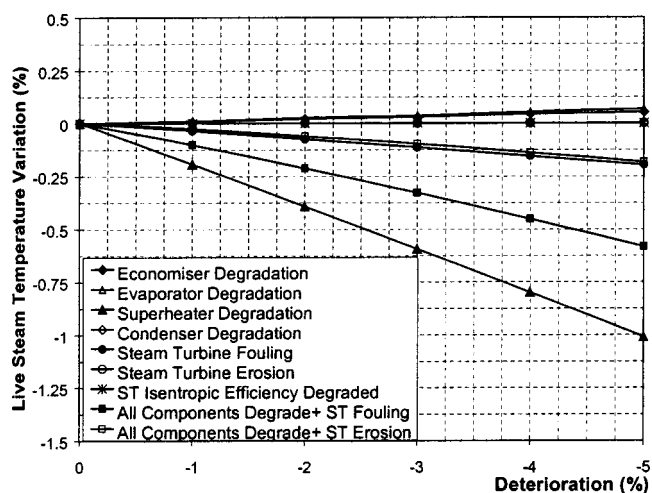


Fig. 7 Live steam temperature variation with steam turbine component deterioration

with steam turbine fouling boosted the inlet pressure to about 6.4%. On the other hand, the steam turbine erosion by 5.0% resulted in about 5.1% reduction in live steam pressure.

As far as live steam temperature concern, the control method during simulation was adjusted in such away that to keep the temperature at the steam turbine inlet not higher than the DP temperature. This is done due to the limitations of mechanical properties of the steam turbine blades. As Fig. 7 shows, the change in live steam temperature was not as high as the change in live steam pressure. The highest change in this case was with superheater degradation, which was -1.1% approximately with 5.0% degradation.

Combined Cycle Degradation Simulation Results

The most important performance deterioration simulation results of combined cycle gas turbine are shown in Figs. 8 through 10. It is an obvious result to find out that the CCGT power output would follow the steam turbine power behavior as the gas turbine power was kept constant during this study, except in the case of increased GT back pressure due to HRSG degradation.

Although the deterioration was high with some steam turbine cycle components degradation, e.g., -5.6% with steam turbine erosion, the GT power output which measures approximately for 2/3's of the CCGT power output did compensate for this reduction in power output.

As Fig. 8 shows, the steam turbine isentropic efficiency degradation and ST erosion accompanied by all components degradation resulted in the higher amount of degradation. As already mentioned, although steam turbine isentropic efficiency degradation by 5.0% resulted in about 5.6% reduction in ST cycle power output, the same amount of degradation resulted in only 1.8% deterioration in CCGT power output. The same comment is applicable to other components degradation.

As in the case of ST power deterioration (see Fig. 3), the effect of steam turbine fouling on CCGT power output (for the same reason mentioned above) was less than the effect of steam turbine erosion.

As Eq. (2) below shows, the combined cycle efficiency is a function of gas turbine cycle efficiency, HRSG efficiency, and steam cycle efficiency, [4].

$$\eta_{CC} = \eta_{GT} + (1 - \eta_{GT}) \cdot \eta_{HRSG} \cdot \eta_{SC} \quad (2)$$

As per this equation, given that the GT efficiency is constant as mentioned above, any decrease in steam turbine efficiency will give its effect directly on CCGT plant efficiency. Even though, as Fig. 10 shows, there was an increase in HRSG efficiency with

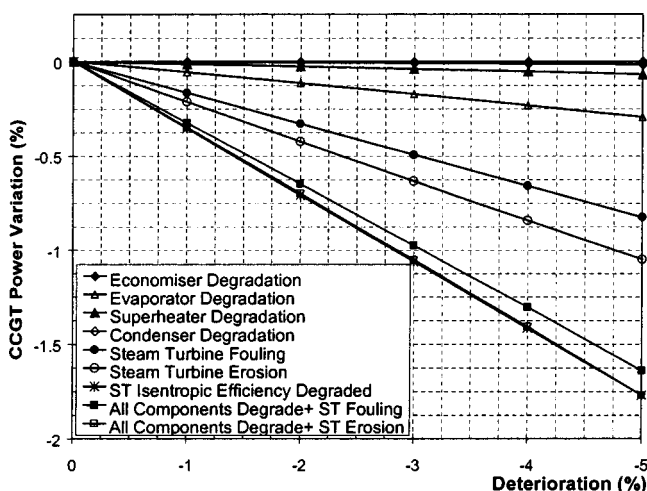


Fig. 8 CCGT power variation with steam turbine component deterioration

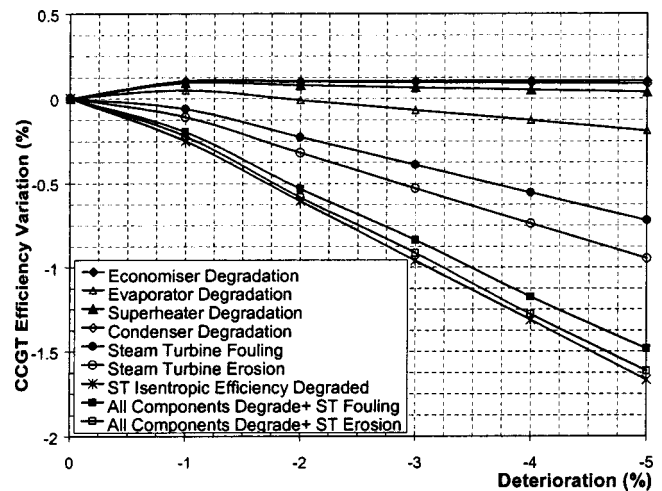


Fig. 9 CCGT efficiency variation with steam turbine component deterioration

some cases of ST component degradation (e.g., all components degrade plus ST erosion), still the effect of decreased ST efficiency for the same fault (see Fig. 4) has a higher effect on CCGT efficiency.

As Fig. 9 shows, the combined cycle efficiency has fallen with all types of ST unit component degradations. The effect of any of the heat exchangers of the HRSG and condenser degradations on CCGT plant efficiency, as can be seen from this figure, is almost negligible. One small exception can be made here where the evaporator degradation by 5.0% led to decrease the CCGT efficiency by about 0.2%.

The stack temperature by itself as a performance parameter is only a measure for the amount of heat extracted from the stream of the GT exhaust gas. It also helps in finding out how far is the HRSG efficient in converting the energy available in the GT exhaust gases to a useful energy. As Fig. 11 shows, the degradation of the combination of all components with ST fouling resulted in the highest increase in stack temperature.

Conclusion

The obtained back-pressure simulation results are in agreement with was found in the open literature.

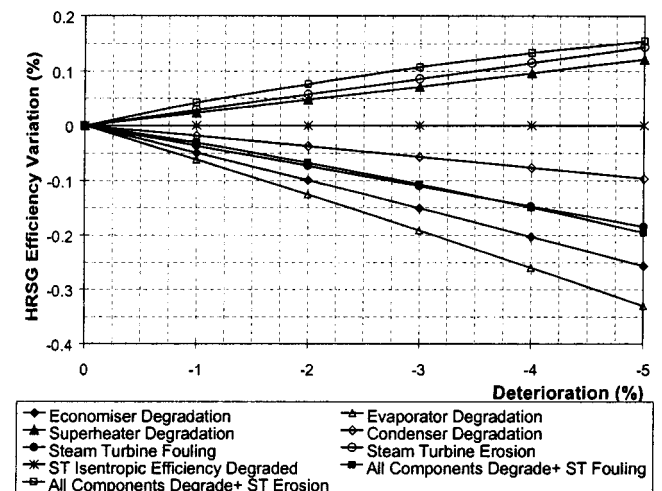


Fig. 10 HRSG efficiency variation with steam turbine component deterioration

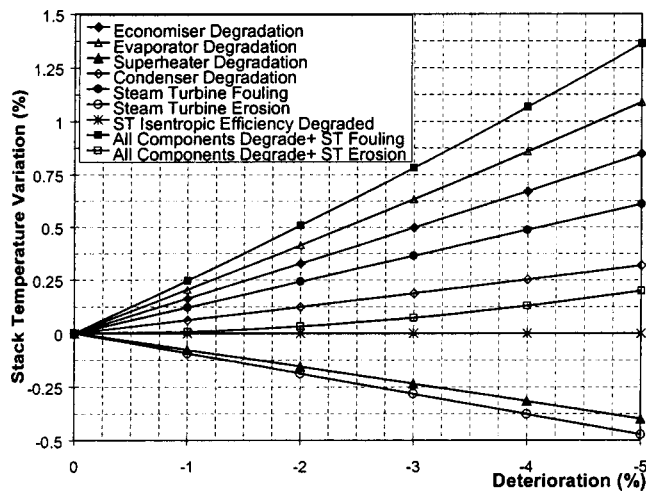


Fig. 11 Stack temperature variation with steam turbine component deterioration

The simulation results obtained from the developed code are in agreement with what the authors put their hands on in the open literature. This gives a sensible confidence in the code developed at Cranfield University, which was the main source of the results in this paper.

Among the three components of HRSG, evaporator degradation is the utmost effecting fault on steam turbine cycle.

The steam turbine isentropic efficiency as a performance parameter has the uppermost effect on steam turbine cycle power and efficiency.

The effect of HRSG and condenser degradations on steam cycle and hence on CCGT plants performance is very low compared to the steam turbine unit components degradation.

Nomenclature

atm = atmospheric pressure
 CCGT = combined cycle gas turbine plant
 DP = design point
 GT = gas turbine
 ST = steam turbine
 Γ = nondimensional mass flow
 η = efficiency
 U = heat transfer coefficient

Subscripts

CC = combined cycle
 GT = gas turbine
 ST = steam turbine
 HRSG = heat recovery steam generator
 i = Inlet
 SC = steam cycle

References

- [1] Saunders, E. A. D., 1988, *Heat Exchangers: Selection, Design and Construction*, Longman Group UK Limited, London.
- [2] Von Nostrand, W. L., et al., 1981, *Fouling of Heat Transfer Equipment*, Hemisphere, Washington, DC.
- [3] Seikan Ishigai, 1999, *Steam Power Engineering: Thermal and Hydraulic Design Principles*, Cambridge University Press, Cambridge, UK.
- [4] Zwebek, A. I., and Pilidis, P., 2001, "Degradation Effects on Combined Cycle Power Plant Performance, Part 1: Gas Turbine Cycle Component Degradation Effects," ASME Paper No. 2001-GT-388.
- [5] Li, Kam W., Priddy, and A. Paul, 1985, *Power Plant System Design*, John Wiley and Sons, New York.
- [6] Sanders, P. William, 1986, *Turbine Steam Path Performance*, Turbomachinery International Publications: Monograph Series, Series No. S.1.

Revisiting the Inhibition of Vanadium-Induced Hot Corrosion in Gas Turbines

E. Rocca

e-mail: emmanuel.rocca@lcsn.uhp-nancy.fr

P. Steinmetz

Université Henri Poincaré–Nancy I,
Laboratoire de Chimie du Solide Minéral,
UMR CNRS 7555,
B. P. 239,
54506 Vandœuvre-Lès-Nancy, France

M. Moliere

GE Energy Products–Europe,
1, Avenue Charles Bohn,
B. P. 379,
90007 Belfort, France
e-mail: michel.molier@ps.ge.com

Since the 1970s, nothing substantially new has been published in the gas turbine community about the hot corrosion by vanadium and its inhibition, after the “inhibition orthodoxy” based on the formation of magnesium vanadate, was established. However, the experience acquired since the late 1980s with heavy-duty gas turbines burning ash-forming fuels in southern China, shows that the combustion of very contaminated fuels does not entail corrosion nor abundant ash-deposit on gas turbines buckets. Analyses of deposits collected from gas turbines fired with these crude oils showed that the ash-deposit contains a large amount of nickel. These new facts led to revisit the role played by nickel and envisage its possible inhibiting action against the vanadium-induced hot corrosion. A thorough review of the literature on the vanadium-induced corrosion have been carried out, and the study of the nickel effects with respect to magnesium effects on the ash deposit have been performed. Results show that nickel presents an interesting way to substitute magnesium for the inhibition of vanadium-induced hot corrosion. The advantages of nickel with respect to magnesium are to be efficient at low Ni/V ratio, to produce less abundant, less adherent ash and to act, to some extent, as a self-cleaning agent for the blades of the turbine. [DOI: 10.1115/1.1456095]

Introduction

Various types of hot corrosion damages caused by vanadium on ferrous and nonferrous alloys have been encountered in all types of combustion equipment burning low-grade fuel oils. This includes boilers, furnaces, diesel engines, and gas turbines. Indeed, petroleum crude and fuel oils can contain up to 4 percent sulfur, 500 ppm vanadium and several tens of ppm of sodium ([1,2]). Ash particles that result from their combustion consist of complex mixtures of vanadium pentoxide (V_2O_5) and sodium sulfate (Na_2SO_4) that exhibit low melting points down to 500°C. These compounds partly deposit on the hot gas path of the thermal equipment (walls, tubes; piston/cylinder; turbine combustors and buckets). The resulting film acts as an ion-electrolyte and is conducive to fast attacks of the substrate material by the oxidizing agents present in combustion gas and in the molten salt film.

Due to their economic impact, these forms of hot corrosion have been the subject of numerous investigations, tests and experiments, especially since the early 1960s when the use of heavy petroleum cuts became intensified in the transportation, heating and power generation sectors ([3–5]). However, despite the mass of industrial/experimental data accumulated in this field, the corrosion mechanisms underlying such damages remain very complex and difficult to control.

The corrosive effects of molten salts in gas turbines or other combustion equipment have been extensively studied. When the salt films contain vanadium, Bornstein et al. [6] proposed that the driving force of the corrosion is the acidic dissolution of the protective oxide scale, because V_2O_5 (like MoO_3 or WO_3) is a strongly acidic oxide. More recently, the oxide solubilities measurements carried out by Hwang [7] and R. A. Rapp [8] showed that the addition of vanadium in such molten salts causes a significant increase of the acidic solubility of the protective oxides. The authors suggested that this behavior results from the com-

plexing effect of oxide ions by vanadate compound, which leads to the formation of fused metal orthovanadate like $AlVO_4$ ($T_f = 695^\circ\text{C}$ [9]) or $CrVO_4$ ($T_f = 810^\circ\text{C}$ [9]). The reaction of dissolution of trivalent oxide can be described as follows: $M_2O_3 + V_2O_5 \rightarrow 2 MVO_4$.

Besides, many authors remarked that the reduction of the vanadium(V) is predominant ([10–12]), and the influence of oxygen pressure is negligible in the corrosion process ([13]).

Consequently, the presence of vanadium in ash-deposit containing sulfate seem to have a double negative effect:

- an increase of oxidising conditions by the redox couple $V(V)/V(IV)$, which facilitates the oxide ion transport across the salt film and so the metal oxidation.
- an increase of acidic oxide dissolution by the complexing effect of vanadate compound.

The usual defense against hot corrosion is the addition of some metals in the fuel as corrosion inhibitors, the concentration of which must be adjusted in function of the level of contaminants in the fuel ([14,15]). For vanadium contaminants, the principle of the hot corrosion inhibition is to trap the corrosive and fused elements in a refractory compound with high melting point; then, the amount of corrosive molten salts on metallic pieces is dramatically reduced. Another inhibiting way, less used, is to dilute the corrosive molten phase in a large excess of an inert refractory compound. Nevertheless, the addition of inhibiting elements must form an easily washable deposit, neither too hard, nor too adherent on the metallic parts of gas turbine, to minimize excessive losses in energy performance.

The most known inhibitor of vanadium-induced hot corrosion is magnesium, which forms MgO in the flame and reacts with V_2O_5 to form a refractory magnesium vanadate $Mg_3V_2O_8$. The inhibiting character of magnesium in gas turbine has been extensively studied by May et al. in the 1970s ([16–18]). According to these works, the magnesium concentration must be adjusted beyond the stoichiometry ratio corresponding to the $Mg_3V_2O_8$ compound. In practice, a minimum ratio of $Mg/V = 3$ is necessary to obtain the maximum inhibiting efficiency. This high ratio, compared to the stoichiometric one, can be explained by several facts:

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, New Orleans, LA, June 4–7, 2001; Paper 01-GT-005. Manuscript received by IGTI, December 2000, final revision, March 2001. Associate Editor: R. Natole.

- A part of MgO is sulfated into MgSO_4 which does not react with V_2O_5 . Indeed, the influence of the SO_2/SO_3 pressure on the inhibiting efficiency of magnesium was noted by Rhys-Jones et al. [19], who has assigned the reduction of the inhibiting effect to the partial sulfatation of MgO.
- The reaction of MgO with V_2O_5 is not complete and is thermodynamically equilibrated.
- The presence of sodium can disturb the formation of $\text{Mg}_3\text{V}_2\text{O}_8$, because of the formation of ternary eutectic phases $\text{MgO}/\text{V}_2\text{O}_5/\text{Na}_2\text{O}$ according to the thermodynamic diagram ([20]). Moreover, May et al. [16] remarked that the optimum ratio must increase with sodium content or Na/V ratio in the fuel. While this inhibition mechanism is a simple one, few things are known as to the effect of additional metals on the inhibition efficiency of magnesium.
- May et al. [21] also studied calcium-based inhibitors, but this element promotes the formation of abundant, compact and hard deposits, which are difficult to wash on blade, because CaO is easily sulfated. Moreover, CaSO_4 is poorly water soluble contrary to MgSO_4 .
- The other element used as inhibitor in gas turbine is chromium. Practically, it is considered as inhibitor of sodium-induced hot corrosion, but its effect and the inhibiting mechanism are not well known. Lowell et al. [22] have tested in a burner-rig a chromium-based commercial inhibitor, intended to reduce sodium-induced hot corrosion, and have noted only a small effect on corrosion. Lee et al. [23] have studied a mixed inhibitor Cr-Mg ($\text{Mg}/\text{V}=3$ and $\text{Cr}/\text{Na}=4$) in a burner-rig and observed a suppression of corrosion by molten salts. The analysis of ash-deposit have shown a large proportion of the very refractory $\text{Mg}_2\text{Cr}_2\text{O}_4$ compound, that may dilute the corrosive molten salt, and acts as an abrasive, removing the ash-deposit on the blades surface. In the point of theoretical view, sodium chromate acts as cicatrizing for the oxide layer in molten sodium sulfate according to Otsuka and Rapp [24]. Nevertheless, in contact with vanadium oxide, Cr_2O_3 diminishes the melting temperature of salts and increases the volume of these corrosive phases due to the low melting point of CrVO_4 , according to the thermodynamic data of Kerby et al. [20]. In conclusion, the effect of chromium is not fully elucidated, although some industrial chromium-based inhibitors seem to be successfully marketed ([25]).
- Other elements like silicium or aluminum have been tested as hot corrosion inhibitors in gas turbines. According to May et al. [26], an important Si/V ratio is necessary to develop a fair inhibiting effect, but the mechanism of inhibition is not obvious. Nevertheless, the addition of aluminum and silicium to a magnesium inhibitor with a suitable ratio can efficiently prevent hot corrosion-induced by the presence of sodium ([27]). In burner rig tests., the $(\text{Mg}-3\text{Al}-4\text{Si})/\text{V}=3$ formulation seems to be efficient for a ratio of $\text{Na}/\text{V}=0.13$ at 900°C ([28]). The mechanism may be attributed to the formation of alumino-silicate refractory compounds which can trap sodium cations. Furthermore, some authors noted a synergistic effect of Cr-Si based inhibitors on sodium-induced hot corrosion ([24,29]).
- Recently, some experience was acquired since the late 1980s with heavy-duty gas turbines burning vanadium-contaminated ash-forming fuels in southern China. Ash-deposit, collected from these machines and analyzed by X fluorescence (Table 1), has the particularity to contain a large amount of nickel. In addition, it was noticed that, despite a large concentration of calcium (inducing hard and abundant deposit) the deposit rate was slow and easily washable in practice.
- In this paper, we have studied the effect of nickel on the sodium-vanadium corrosive melts. In a first step, analysis and thermochemical studies of nickel-vanadium-sodium combinations were carried out. Then, lab-scale corrosion studies

Table 1 Composition by X-ray fluorescence of three samples removed from turbine blades in a gas turbine installed in southern China

(wt.%)	Sample 1	Sample 2	Sample 3
SO_3	36.92	37.26	36.54
CaO	24.2	25.08	24.18
NiO	16.83	15.14	19.11
Fe_2O_3	2.67	3.22	3.65
MgO	2.85	3.97	2.01
V_2O_5	0.96	0.99	1.38
TiO_2	0.85	0.45	0.7
$\text{K}_2\text{O} + \text{Na}_2\text{O}$	0.72	0.59	0.61
Loss on fire	9	9.58	7.78
Impurities (Co, Cr, Ti, Si, Al, P)	3.2	1.5	1
TOTAL	99.2	99.78	99.96

were performed on turbine superalloys by sodium sulfate/vanadate melts, and the inhibiting efficiency of nickel was evaluated. Finally, burner rig tests were realized to evaluate the ash-deposit rates along with the corrosion in more realistic conditions.

Experimental

Thermochemical Tests. The experiments of thermochemical study of oxides were conducted from the reagent mixtures that were ground and pressed, in tube furnaces at 700°C in air for three days. NiO, MgO, Na_2SO_4 , V_2O_5 were obtained from Aldrich. The products of reaction were determined by X-ray diffraction and metallographic observations. The metallographic cross sections were prepared by embedding the sample in epoxy resin under vacuum followed by a polishing in nonaqueous solvent, and were analyzed by energy dispersive spectroscopy in a scanning electron microscope. The melting temperature of some pure phases was measured by differential thermal analyses to verify thermodynamic data.

Lab-Scale Corrosion Tests. The corrosion tests in laboratory were performed in a vertical tube furnaces at 850°C and 700°C for 100h using parallelepiped metallic samples ($10 \times 10 \times 2$ mm) suspended with platinum wire in air containing 150ppm SO_2 at 3 l.h^{-1} . Before the tests, the metallic samples (GTD111 alloy) were covered by an oxide deposit of 5 mg.cm^{-2} to simulate the corrosive-ash deposit, by spraying an aqueous solution containing corrosive and inhibitive elements on the metallic sample, previously heated at 400°C . The sprayed aqueous solutions were prepared to adjust a constant content of sodium, sulfur, and vanadium ($\text{V}_2\text{O}_5/\text{Na}_2\text{SO}_4=70/30$ in mass) and the desired values of the Ni/V or Mg/V ratio (in metal mass) in the deposit. The metallographic cross sections of the oxidized sample were observed by SEM and analyzed by EDS. The sprayed solutions were prepared from $\text{Ni}(\text{NO}_3)_2 \cdot 6\text{H}_2\text{O}$, $\text{Mg}(\text{NO}_3)_2 \cdot 6\text{H}_2\text{O}$, Na_2SO_4 dissolved in deionized water and V_2O_5 in suspension.

Burner-Rig Tests. Burner-rig experiments were carried out to investigate the corrosion of superalloys in dynamic conditions of ash-deposition (test duration: 250h). The fuel feeding the burner rig was domestic gasoil with 0.2wt.% of sulfur. An aqueous solu-

tion containing adjusted concentrations of sodium, vanadium and nickel or magnesium was sprayed in the combustion gas just after the burner. Within the burner-rig tests, the sprayed solution was set up by soluble salts in aqueous solution (VOSO_4 , $5\text{H}_2\text{O}$, NaCl , and $\text{Ni}(\text{NO}_3)_2 \cdot 6\text{H}_2\text{O}$ or $\text{Mg}(\text{NO}_3)_2 \cdot 6\text{H}_2\text{O}$) to avoid problems like the formation of aqueous suspension during spraying. The alloy samples were GTD111 and GTD222 rods with 8 mm diameter and 8 cm long, placed in ceramic support which rotates in gas combustion flow. The samples were maintained at a temperature of 850°C during all the tests by controlling the temperature of the furnace, and were quenched in air during 7 min every 12 hours to simulate thermal shocks. The different parameters of the burner rig were adjusted to obtain an ash-deposition rate of about $5\text{ mg}\cdot\text{cm}^{-2}$ for 100 h on the samples. Inert reference samples were placed on the ceramic support to control the deposition rate every 75 h. Considering all the parameters of the burner rig, the fuel and the solution sprayed in the combustion gas simulate the combustion of a crude oil with a composition of 170 ppm V, 4 ppm Na, and 0.2 wt.%S.

After the burner-rig tests, the samples were analyzed by metallographic observations and by the measurement of weight gain.

Results and Discussion

The trapping of vanadium metal in a refractory compound is the simplest system to inhibit the vanadium-induced hot corrosion. Vanadate, VO_4^{3-} that is formed by oxidative combustion of vanadium organometallics of the fuel in combustion chamber, can react with divalent cations M^{2+} to trap vanadium into a refractory compound of $\text{M}_3\text{V}_2\text{O}_8$ type. The efficiency of the inhibiting action of M^{2+} cation is directly linked to the thermodynamic and kinetic properties of the trapping reaction, which depends on the type of cation and its concentration in the fuel.

1 Thermodynamic Studies of $\text{NiO}/\text{V}_2\text{O}_5$ System. Results show that the reaction between NiO and V_2O_5 produces very quantitatively nickel vanadate $\text{Ni}_3\text{V}_2\text{O}_8$ in sodium sulfate melt under air + 150ppm SO_3 , which is in accordance with the thermodynamic diagram $\text{NiO}-\text{V}_2\text{O}_5$ (Figs. 1 and 2). Indeed, the formation of $\text{Ni}_3\text{V}_2\text{O}_8$ is complete as soon as the Ni/V ratio ($=1.74$) reaches the stoichiometry of $\text{Ni}_3\text{V}_2\text{O}_8$ compound (Fig. 1(b)), and the following reaction is then total: $3\text{NiO} + \text{V}_2\text{O}_5 \rightarrow \text{Ni}_3\text{V}_2\text{O}_8$. If the ratio is lower than 1.74, an excess of V_2O_5 is present in the melt, which reacts with Na_2SO_4 to form molten NaVO_3 at the grain boundaries of sodium sulfate (Fig. 1(a)). According to Hwang et al. [7], sodium sulfate and vanadium oxide reacts as follow: $\text{Na}_2\text{SO}_4 + \text{V}_2\text{O}_5 \rightarrow 2\text{NaVO}_3 + \text{SO}_3$. Nevertheless, XRD and SEM analyses show that the refractory compound $\text{Ni}_3\text{V}_2\text{O}_8$ is even so formed quantitatively in these conditions.

In all these experiments in presence of 150 ppm of SO_2 , the formation of NiSO_4 has not been observed because of kinetic limitation. Indeed, the reaction between NiO and V_2O_5 seems to be very fast and the nickel is immediately incorporated into the vanadate compound.

Besides, pure $\text{Ni}_3\text{V}_2\text{O}_8$ was synthesized by reaction of a stoichiometric melt of NiO and V_2O_5 at 1000°C in air, and its melting point was measured by ATD experiment. The experimental value of 1310°C is in accordance with thermodynamic data, and is higher than the one of $\text{Mg}_3\text{V}_2\text{O}_8$ (1200°C) ([9]).

2 Lab-Scale Corrosion Tests. To test the effect of nickel addition on the hot corrosion of superalloys and to compare it to magnesium addition, lab-scale corrosion tests were performed at 850°C and 700°C with superalloys samples covered by a synthetic and corrosive ash-deposit. In these tests, a Ni/V ratio equal to 2.25 was used, in comparison to the classical Mg/V ratio of 3. The choice of $\text{Ni}/\text{V}=2.25$ allows to add a low excess of nickel in the melts, in comparison with the stoichiometry ratio of 1.74 corresponding to the formation of $\text{Ni}_3\text{V}_2\text{O}_8$ compound.

At 850°C without nickel and magnesium in the deposit, the corrosion of GTD111 is very important as it can be seen in Fig. 3.

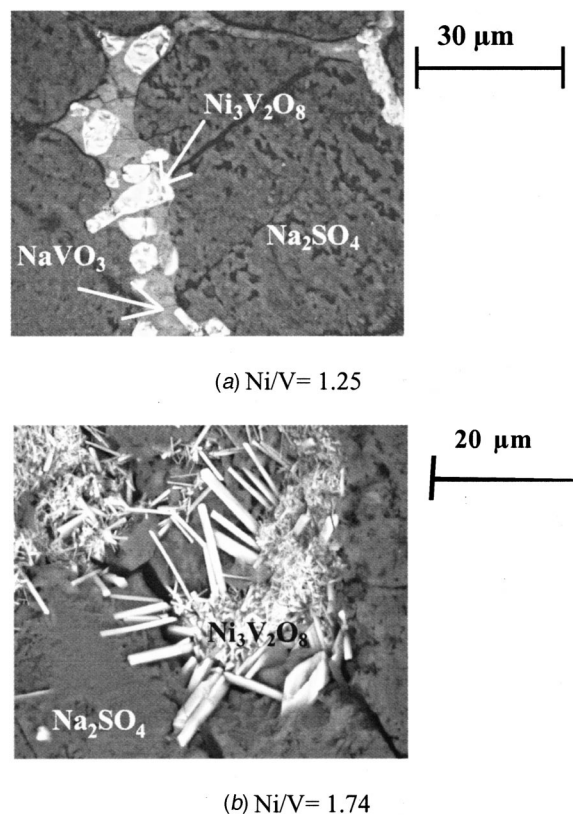


Fig. 1 Metallographic cross-sections of $\text{NiO}/\text{V}_2\text{O}_5$ melts in Na_2SO_4 (70 wt.%) after three days at 700°C in air + 150 ppm SO_2

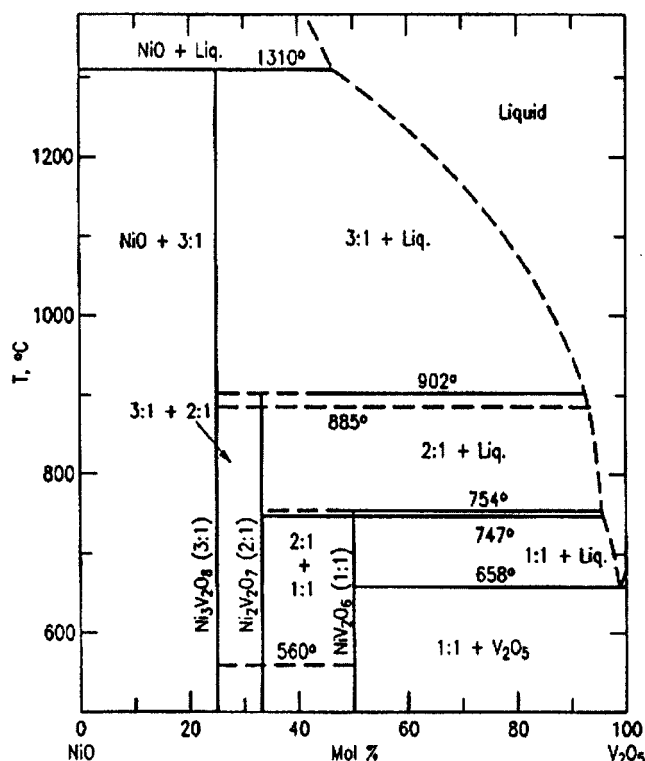


Fig. 2 $\text{NiO}-\text{V}_2\text{O}_5$ thermodynamic diagram ([9])

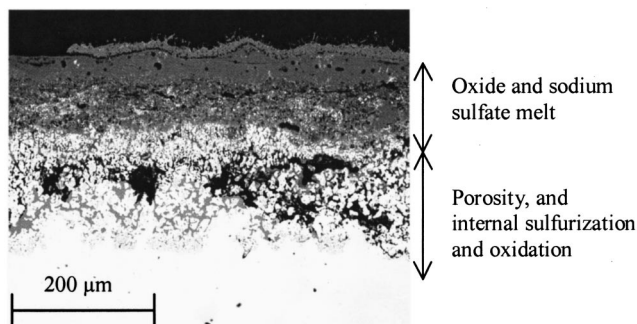
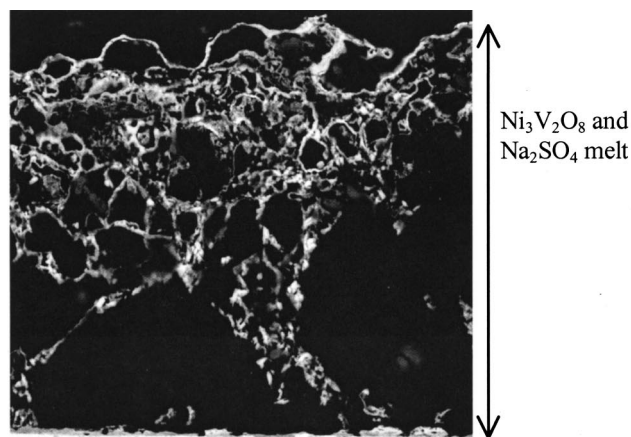
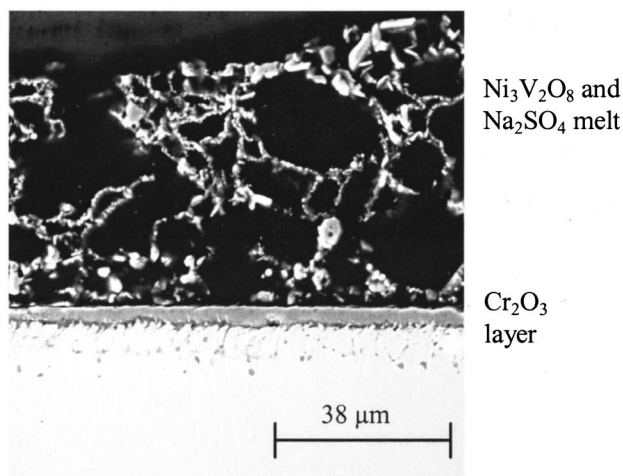


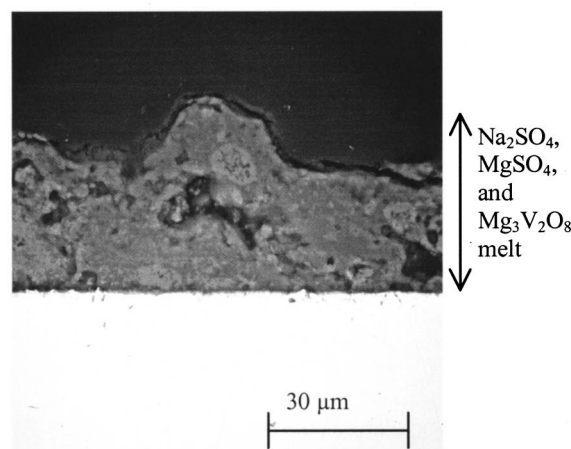
Fig. 3 Metallographic cross section of GTD111 covered with $\text{Na}_2\text{SO}_4/\text{V}_2\text{O}_5$ melt and oxidized at 850°C in air+150 ppm SO_2 for 100 h



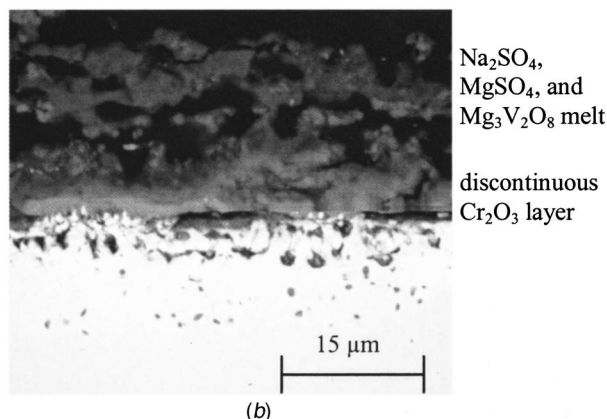
(a)



(a)



(b)



(b)

Fig. 4 Metallographic cross section of GTD111 covered with $\text{Na}_2\text{SO}_4/\text{V}_2\text{O}_5$ inhibited with nickel ($\text{Ni}/\text{V}=2.25$) (a), and inhibited with magnesium ($\text{Mg}/\text{V}=3$) (b). The samples were oxidized at 850°C in air+150 ppm SO_2 for 100 h.

The constituting elements of the GTD111 alloy are stuck in the ash-deposit, which induces the formation of porosity in the bulk of metallic pieces. Indeed, nickel and chromium are dissolved into the sodium sulfate-vanadate melt, and the other elements are transformed into oxide inside the ash-deposit.

Furthermore, inside the alloy, we can notice a sulfidation and an

Fig. 5 Metallographic cross section of GTD111 covered with $\text{Na}_2\text{SO}_4/\text{V}_2\text{O}_5$ inhibited with nickel ($\text{Ni}/\text{V}=2.25$) (a), and inhibited with magnesium ($\text{Mg}/\text{V}=3$) (b). The samples were oxidized at 700°C in air+150 ppm SO_2 for 100 h.

oxidation of some elements like chromium, and a dechromization around these penetrations, in accordance with many authors.

In opposite, when nickel ($\text{Ni}/\text{V}=2.25$) or magnesium ($\text{Mg}/\text{V}=3$) is added to the ash-deposit, the corrosion of the alloy is drastically reduced (Fig. 4). No porosity is formed in the alloy bulk, and we can only observe a slight internal oxidation and sulfidation. On the alloy surface, the presence of a passive chromine layer proves that the hot corrosion of the metal is very low, and is replaced by a classical phenomenon of oxidation by gases. Nevertheless, we can remark that the Cr_2O_3 layer is less compact and more discontinuous on the sample inhibited with magnesium than on the one inhibited with nickel.

Moreover, the structure of the ash-deposit containing magnesium is much less porous and crumbly than the one containing nickel. This fact can be due to the excess of magnesium, which produces MgSO_4 and acts as a binding agent in the structure of the deposit. It is noteworthy that the alveolar structure of the nickel-containing deposit is very interesting for the purpose of washing the gas turbine.

At 700°C , the difference of the compactness between the two deposits is much greater than at 850°C , because the rate of magnesium sulfatation is more important (Fig. 5). Indeed, the reaction between MgO and V_2O_5 to form $\text{Mg}_3\text{V}_2\text{O}_8$ is in competition with

Table 2 Weight gain rate for the GTD111 sample and inert samples in burner rig in different conditions

Combustion conditions	Weight gain of GTD111 samples (mg/cm ² /100h)	Calculated weight gain due to corrosion (mg/cm ² /100h)	Weight gain of inert samples (mg/cm ² /100h)
Non-inhibited	5.25	0.56	4.69
Mg inhibited (Mg/V=3)	2.01	0.33	1.68
Ni inhibited (Ni/V=2.25)	0.84	0.14	0.70

the transformation of MgO into MgSO₄. This observation is in agreement with some authors ([19]) and can explain the need in practice to overdose magnesium with respect to vanadium.

Nevertheless, for both magnesium and nickel at 700°C, the vanadium-induced hot corrosion is inhibited with a great efficiency, and we can remark the formation of a thin Cr₂O₃ layer after 100 h of oxidation.

These corrosion tests show that nickel is at least as efficient as magnesium to inhibit vanadium-induced hot corrosion, and has the advantage to form a very crumbly ash-deposit and to be required at a lower concentration than magnesium.

3 Burner-Rig Tests. Burner-rig tests were carried out to test this nickel-based inhibitor in dynamic conditions of ash-deposit. After 250 h of combustion, the net weight gain rate, including both the ash-deposit rate on the piece and the corrosion rate of the alloy, were constant and reproducible for GTD111 alloy in the given conditions. The ash deposition rate is evaluated by the weight gain of inert samples. Table 2 gives the values in the three cases.

It is noteworthy that the weight gain of the Ni-inhibited GTD111 samples is six times lower than for the noninhibited samples, and the Mg-inhibited samples has a weight gain twice than that of the Ni inhibited one. Moreover, the ash-deposit on the Ni inhibited samples is very crumbly and brittle, and XRD analyses have confirmed that it is mainly composed by Ni₃V₂O₈ and a very low excess of NiO. In opposite, the ash-deposit on Mg-inhibited samples are composed by the refractory vanadate Mg₃V₂O₈ and a large amount of MgO, which explains the weight gain higher than the one of Ni-inhibited samples.

A specific cleaning procedure for the removal of nickel ash traces that can nevertheless accumulate in the gas turbine hot gas path, has been developed ([30]).

Conclusion

Alcaline-earth metals, which combine with vanadium to form high melting point salts, have been early identified as inhibitors of the vanadium corrosion. Especially, in the gas turbine industry, the use of magnesium, which moreover forms washable deposits has become the “inhibition orthodoxy” in the 1970s. This type of inhibitor has proven quite efficient over several decades of operation in heavy fuelled gas turbines. Nevertheless, the inhibition of vanadium-induced hot corrosion by magnesium requires frequent cleaning of gas turbine blades and limits the operating temperature of gas turbines.

The efficiency of nickel as novel inhibitor of vanadium-induced hot corrosion has been demonstrated. The potential advantages of nickel with respect to magnesium are as follows:

- The melting point of the formed refractory vanadate (T_fNi₃V₂O₈ = 1310°C) is higher than the one of magnesium (T_fMg₃V₂O₈ = 1200°C).
- The reaction of NiO with V₂O₅ is very fast and complete.

- The formed ash deposits are not very abundant and very crumbly. Nickel seems to be a self-cleaning agent.
- The parasite reaction of sulfatation of NiO is much less important than the one of MgO; consequently, an overdosage in nickel is not necessary.

So, nickel represents an interesting alternative to substitute magnesium for the inhibition of hot vanadium corrosion in gas turbine. It should allow an increase in the operating temperature, and less frequent washes of the turbine blades.

The overall concept of using nickel as an industrial inhibitor has been patented ([31]).

Acknowledgments

The authors wish to thank the “Laboratoire des Refractaires et des Matériaux Céramiques” (LRMC, Nancy, France) who supplied the special burner rig tests at their facility which proved very flexible. WEB site: www.lrmc.fr or e-mail: info@lrmc.fr.

References

- [1] Luthra, K. L., and Spacil, H. S., 1982, “Impurity Deposits in Gas Turbines From Fuels Containing Sodium and Vanadium,” *J. Electrochem. Soc.*, **129**, pp. 649–656.
- [2] Kofstad, P., 1988, *High Temperature Corrosion*, Elsevier, London.
- [3] Wickert, K., 1961, “High Temperature Corrosion of Oil-Fired Equipment,” *Oelfeuerung*, **6**, pp. 602–610.
- [4] Winkler, F., 1978, “Relationship of Fuel Quality to Marine Boilers and Diesel Engines Performances,” *Soc. Marine. Port Eng.*, Jan.
- [5] Geisler, H., 1962, “Oelgefeuerte Dampfkesselanlagen und ihre Probleme,” *Energie*, Jun-Sept., Mar. 1963.
- [6] Bornstein, N. S., DeCrescente, M. A., and Roth, H. A., 1973, “The Relationship Between Relative Oxide Ion Content of Na₂SO₄, the Presence of Liquid Metal Oxides and Sulfidation Attack,” *Metallurgical Transactions*, **4**, pp. 1799–1810.
- [7] Hwang, Y.-S., and Rapp, R. A., 1989, “Thermochemistry and Solubilities of Oxides in Sodium Sulfate-Vanadate Solutions,” *Corrosion*, **45**, pp. 933–937.
- [8] Rapp, R.-A., and Zhang, Y.-S., 1994, “Hot Corrosion of Materials: Fundamentals Studies,” *Dec. J.O.M.*, pp. 47–55.
- [9] Levin, E. M., Robbins, C. R., and McMurdie, H. F., 1964, “Phase Diagrams for Ceramists,” compiled by the National Bureau of Standards, The American Ceramic Society, Columbus, OH.
- [10] Seirstein, M., and Kofstad, P., 1987, “Sodium Vanadate-Induced Corrosion of Nickel and MCrAlY Coatings on Inconel 600,” *Mater. Sci. Technol.*, pp. 576–583.
- [11] Otero, E., Pardo, A., Perez, F., Alvarez, J. F., and Utrilla, M. V., 1997, “A High Temperature Corrosion Kinetic Study of HK-40 Superalloy Surface Treated, in Contact With Eutectic Mixture 82%K₂S₂O₇-18%V₂O₅,” *Corros. Sci.*, **39**, pp. 133–145.
- [12] Singh, I. B., 1998, “V₂O₅ Accelerated Hot Corrosion of Nickel in Alkali Sulphates Melt,” *Ind. J. of Chem. Technol.*, **5**, pp. 163–165.
- [13] Otero, E., Pardo, A., Hemaes, J., and Perez, F. J., 1992, “The Corrosion Kinetic of Some Superalloy (at 1000K) in Molten Eutectic Mixture 60%V₂O₅-40%Na₂SO₄. The Influence of the Oxygen and Carbon Residues,” *Corros. Sci.*, **33**, pp. 1747–1757.
- [14] Molliere, M., 1993, “Experience With Gas Turbines Burning Ash-Forming Fuels,” *GE EPE Library, Tech. Rev.*, **11**, pp. 47–61.
- [15] Molliere, M., 1998, “Volatile, Low Lubricity Fuels in Gas Turbine Plants: A Review of Main Fuel Options and Their Respective Merits,” *ASME Paper 98-GT-231*.
- [16] May, W. R., Zetlmeisl, M. J., and Bsharah, L., 1973, “High Temperature Corrosion in Gas Turbines and Steam Boilers by Fuel Impurities. II. The Sodium Sulfate-Magnesium Sulfate-Vanadium Pentoxide System,” *Ind. Eng. Chem. Prod. Res. Dev.*, **12**, No. 2, pp. 140–144.
- [17] May, W. R., Zetlmeisl, M. J., Bsharah, L., and Annand, R. R., 1973, “High Temperature Corrosion in Gas Turbines and Steam Boilers by Fuel Impurities. III. Evaluation of Magnesium as a Corrosion,” *Ind. Eng. Chem. Prod. Res. Dev.*, **12**, No. 2, pp. 145–149.
- [18] May, W. R., Zetlmeisl, M. J., and Annand, R. R., 1976, “High Temperature Corrosion in Gas Turbines and Steam Boilers by Fuel Impurities,” *ASME J. Eng. Gas Turbines Power*, **98**, pp. 506–510.
- [19] Rhys-Jones, T. N., Nicholls, J. R., and Hancock, P., 1983, “Effects of SO₂/SO₃ on the Efficiency With Which MgO Inhibits Vanadic Corrosion in Residual Fuel Fired Gas Turbine,” *Corros. Sci.*, **23**, No. 2, pp. 139–149.
- [20] Kerby, R. C., and Wilson, J. R., 1973, “Solid-Liquid Phase Equilibria for the Ternary Systems V₂O₅-Na₂O-Fe₂O₃, V₂O₅-Na₂O-Cr₂O₃, and V₂O₅-Na₂O-MgO,” *Can. J. Chem. Eng.*, **51**, pp. 1032–1040.
- [21] May, W. R., Zetlmeisl, M. J., Annand, R. R., and Laurence, D. F., 1976, “High Temperature Corrosion in Gas Turbines and Steam Boilers by Fuel Impurities. III. Evaluation of the Effects of Manganese, Calcium and Several Heavy Metals on Corrosion and Slag Formation,” Presented to the Symposium on Heavy Fuel Oil Additives, Apr., New York.

- [22] Lowell, C. E., and Deadmore, D. L., 1978, "Effect of a Chromium-Containing Fuel Additive on Hot Corrosion," *Corros. Sci.*, **18**, pp. 747–763.
- [23] Lee, S. Y., Chamberlin, R. M., and Spengler, C. J., 1978, "Control of Corrosion in Combustion Turbines With Mg-Cr Combined Fuel Additives," presented at the Winter Annual meeting of the ASME, Dec. 10–15, San Francisco, CA.
- [24] Otsuka, N., and Rapp, R. A., 1990, "Effect of Chromate and Vanadate Anions on the Hot Corrosion of Preoxidized Ni by a Thin Fused Na_2SO_4 Film at 900°C," *J. Electrochem. Soc.*, **137**, No. 1.
- [25] J. P. Stalder and P. A. Huber, 2000, "Use of Chromium Containing Fuel Additive to Reduce High Temperature Corrosion of Hot Section Parts" presented at the International Gas Turbine and Aeroengine Congress & Exhibition, May 8–11, Munich, Germany.
- [26] May, W. R., Zetlmeisl, M. J., and Annand, R. R., 1974, "High Temperature Corrosion in gas Turbines and Steam Boilers by Fuel Impurities. III. Evaluation of Magnesium as a Corrosion," *ASME J. Eng. Gas Turbines Power*, **96**, pp. 124–128.
- [27] Zetlmeisl, M. J., Purdy, D. R., and Laurence, D. F., 1979, "A Laboratory Investigation and Regression Analysis of Boiler Deposition as a Function of Na, V, Mg, Si, Al," presented at the Winter Annual Meeting of the Dec. 7–9, ASME, New York.
- [28] May, W. R., Zetlmeisl, M. J., and Annand, R. R., 1976, "High Temperature Corrosion in Gas Turbines and Steam Boilers by Fuel Impurities. VI. Evaluation of Magnesium-Aluminum-Silicon Combinations as a Corrosion Inhibitors," *J. Eng. Power*, **98**, pp. 511–516.
- [29] Zetlmeisl, M. J., Laurence, D. F., and McCarthy, K. J., 1984, "Evaluation of Sulfidation Corrosion Inhibitors by Small Burner Rig Testing," *Mat. Perf.*, June, pp. 41–44.
- [30] Lecuire, J. M., Magri, P., Michel, S., and Hody, V., 2000, "Chemical Removal of $\text{Ni}_3\text{V}_2\text{O}_8\text{-CaSO}_4$ Deposits From Gas Turbine Blade," Final Contract Report, University of Metz, France.
- [31] GE Patent No. 00 07806.

Flashback in Lean Prevaporized Premixed Combustion: Nonswirling Turbulent Pipe Flow Study

O. Schäfer

e-mail: olaf.Schaefer@its.uni-karlsruhe.de

R. Koch

S. Wittig

Institute of Thermal Turbomachinery,
Department of Mechanical Engineering,
University of Karlsruhe,
Kaiserstrasse 12,
76128 Karlsruhe, Germany

A fundamental study has been performed on the upstream flame propagation of a turbulent kerosene flame, stabilized in a confined stagnation flow at atmospheric pressure. Besides temperature and equivalence ratio, mixture properties and fluid dynamic parameters have been varied. The flashback phenomenon is discussed in terms of critical mean velocities and additionally based on detailed LDV data at the outlet of the premixing duct. The largest critical velocities u_c for flashback are found for the "perfectly" premixed case and equivalence ratios close to stoichiometric, which is in accordance with the theory on laminar flame propagation. In the case of a homogeneous mixture, flashback is determined by the velocity distribution at the outlet of the premixing section. In the undisturbed pipe flow the flame propagates through the wall boundary layer. The data for this case are compared with the theory of side-wall quenching in terms of a critical Peclet number and critical velocity gradients at the wall. Both are deduced from the experimental data. Reducing the velocity on the axis forces the flame to propagate through the center at a velocity predicted by correlations on turbulent flame velocity. [DOI: 10.1115/1.1581897]

Introduction

A major challenge to gas turbine engineering today is the substantial reduction of NO_x emissions without compromising CO and UHC, [1,2]. One promising concept for liquid fuel operation is lean prevaporized premixed (LPP) combustion. Unfortunately, the NO_x reduction potential of LPP combustion is impaired by operational disadvantages, i.e., combustion instabilities, susceptibility to auto-ignition, and flashback.

Flashback in LPP combustors can be defined as the fast chemical reaction, accompanied by significant heat release in the premixing section of the combustor, owing to upstream propagation of a flame from the main combustion zone, [3]. For a clear differentiation between flashback and auto-ignition the residence time of the mixture within the premixing duct should be much smaller than the ignition delay.

Most data on flashback have been obtained for premixed, laminar gas flames at room temperature and atmospheric or subatmospheric pressure, [4–6]. A limited number of experiments were also conducted at elevated temperatures, [7–9], and a few in turbulent flow, [7,10]. All those studies are unfortunately on gaseous premixed systems with hydrogen or pure hydrocarbons as fuels and do not include mixedness or heterogeneous effects. The first theoretical considerations on flashback in laminar pipe flow are based on the critical wall velocity gradient concept developed by Lewis and von Elbe [11]. In this concept flashback occurs if the velocity gradient at the wall is less than the ratio of the flame speed to the quenching distance. Putnam and Jensen [5] extended the gradient concept into a dimensionless formulation based on the Peclet number to include gas composition or burner pressure.

For combustion systems of technical scale Plee and Mellor [12] reviewed reported flashback data and classified four mechanisms of upstream flame propagation in LPP systems:

1 auto-ignition upstream of the flame due to the mean residence time in the premixing duct exceeding the fuel ignition delay time,

2 classical flashback due to the normal burning velocity exceeding the flow velocity either in the boundary layer or along the flow axis,

3 flame propagation through reverse flow fields that may occur when the characteristic dimension of a disturbance is of the same order as the distance between the disturbance and the flame holder, and

4 auto-ignition in separated flow regions within the premixing duct that may take place when the dimension of the disturbance is much smaller than the distance between the disturbance and the flame holder.

In many cases flashback and upstream flame propagation were found to be induced by disturbances causing flow separation in the premixing tube. Coats [13] mentioned that in systems which utilize combustion stabilization in mixing layers, i.e., bluff-body flameholders, flashback, and combustion instabilities were observed simultaneously. In those experiments flashback was observed at velocities higher than those expected from classical theory. Several studies in the context of combustor technology for afterburner and hypersonic transportation applications have confirmed the importance of local flow reversal triggered by large-scale vortex dynamics in reacting shear layers with respect to flashback, [14–16]. Coincidence of the natural vortex shedding frequency on a backward facing step with resonance frequency of the combustor can lead to large pressure fluctuations. This causes the flame to lift from the edge of the step and propagate upstream. Similar results have been reported by Marek et al. [17] for a liquid fueled dump combustor operated at pressures up to 0.56 MPa and by Proctor et al. [18] for a propane fueled center-dump combustor at atmospheric pressure. Due to the lack of optical accessibility in high pressure tests, upstream flame stabilization is mostly discovered indirectly by burnt out parts or thermocouple readings. The reasons for upstream flame stabilization have to be inferred from indirect evidence rather than from direct measurements, [19]. Hence no comprehensive data on the flashback phenomenon for

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, New Orleans, LA, June 4–7, 2001; Paper 2001-GT-0053. Manuscript received by IGTI, Dec. 2000, final revision, Mar. 2001. Associate Editor: R. Natole.

LPP systems on engine scale, operating with aircraft fuels premixed at high temperatures and pressures, is presently available in the literature.

The variety of the mentioned mechanisms makes it quite difficult to determine which mechanism or combination of mechanisms is responsible for flashback in each case. In addition, data on the flame speed of liquid fuels at high temperature and pressure, which are required in the theoretical analysis, may often be difficult to obtain or may not even exist.

Our approach to the flashback phenomenon is twofold and in all cases is closely linked to the practical application. The experiment is designed to gradually increase the degree of complexity, starting from a nonswirling pipe flow, adding swirl and finally including effects of combustion oscillations. The analysis of the phenomenon is realized in a similar manner, starting from mean velocities at flashback and going into the details of the velocity distributions at the onset of flashback. The results to be presented here describe the first phase of a fundamental and comprehensive study on the flashback phenomenon.

Experimental Setup and Techniques

Test Rig. The experiment has been designed to realize different types of stabilization of lean premixed flames, e.g., stagnation flow, dump combustor, and swirl flow. Only the premixed stagnation flame in a nonswirling flow will be discussed herein. In this case the primary fuel-air mixture and the secondary air enter the combustor coaxially, impinge on the movable, air cooled stagnation plate and stabilize a planar turbulent flame in the vicinity of the plate. A sketch of the atmospheric test rig is given in Fig. 1. Due to the large ignition delay times at atmospheric conditions, auto-ignition effects do not have to be considered.

Primary and secondary air mass flows are controlled and heated separately. Both mass flows are measured by means of calibrated hot film sensors. Two 15 kW electrical heaters are used to heat the air flows up to a maximum temperature of 973 K. Before entering the main part of the test rig, shown in Fig. 1, the primary air passes a settling chamber. The settling chamber is equipped with static mixing elements and flow straighteners to homogenize the temperature and the velocity distribution. Additionally it is surrounded with heating coils to compensate for heat losses. Kerosene (Jet-A1) is injected via an internally mixing twin-fluid atomizer (Fluidics Instruments). The mass flow rate of the kerosene is calculated from the volume flow rate, metered via a turbine flowmeter, with the density as a linear function of temperature. The atomizing air mass flow is measured with a rotameter. To calculate the equivalence ratio a stoichiometric air to fuel mass ratio of 14.62 is assumed, based on a chemical analysis of the kerosene used. The atomizer produces very small droplets ($SMD < 15 \mu m$) over the whole operating range and permits two discrete spray cone angles (10 deg or 20 deg). The premixing duct consists of two parts, a quartz glass tube and a metal plenum flange. The

glass tube is 0.4 m in length and provides optical access to the flame propagation in the upstream part of the premixing duct. It permits the installation of three static mixing elements (Sulzer Chemtech, SMX) to homogenize the mixture. The glass tube can be heated from outside to provide a flat temperature profile (± 5 K) at the combustor inlet. This is realized at the expense of optical accessibility. Several elements can be inserted in the premixing tube: Perforated plates can be inserted at two locations ($x = -0.09$ and -0.05 m) to change the velocity and turbulence profile at the combustor inlet. Furthermore, a wire gauze can be inserted ($x = -0.12$ m and -0.15 m) to relaminarize the flow and to stabilize a planar laminar flame in front of the stagnation plate. The overall length of the premixing duct is 0.55 m. Within the plenum flange the secondary air is merged with the primary air in a coaxial fashion. The inner and outer diameter of the primary and secondary flow at the inlet of the combustor are 0.041 and 0.058 m, respectively, yielding an area ratio of one. At the combustor inlet the premixing tube is chamfered by 4 deg on the inner and the outer side to obtain a very thin shear layer between the primary and secondary flow. The annular flow, a typical feature of stagnation flame investigations, inhibits the entrainment of hot combustion products that recirculate in the corners of the combustor. Mean flow rate and temperatures of primary and secondary air are equal if no further specification is made in the text. The coaxial jet enters the combustion chamber ($d_c = 0.15$ m) and impinges on a movable stagnation plate ($d_p = 0.12$ m). The stagnation plate is water cooled except for the head plate. The head plate is air cooled to avoid a large heat flux from the turbulent flame to the plate.

Measurement Techniques. Detailed information on the velocity distribution at the inlet to the combustor is acquired by means of LDV measurements. The LDV system consists of an argon-ion-laser, operated at the power of 1.5 W, and a four beam two color Dantec Fiber Flow system in backscatter mode. The measurement volume is 0.219×4.611 mm in diameter and length, respectively. Two burst spectrum analyzers (BSA 57N20) are used to process the data. $CaCO_3$ particles ($d < 1 \mu m$) are equally added to both flows for seeding. The orientation of the beams permits the measurement of the velocity components in the x and z -directions (cp. Fig. 1). Velocity distributions at the inlet to the combustor are measured in the yz -plane at $x = 0.003$ m through a small opening of 0.01 m between the plenum and the glass cylinder. For the fundamental analysis of flashback the velocity distribution at the inlet to the combustor should be measured at the incident of flashback. As this is not possible with a point measuring technique like LDV, the velocity distributions have to be acquired for mean velocities very close to flashback.

Measurement of the pressure oscillations within the combustion chamber ensures the detection of flashback, free of nonstationary effects associated with combustion instabilities. The pressure probe consists of a condenser microphone (Bruel & Kjaer Type

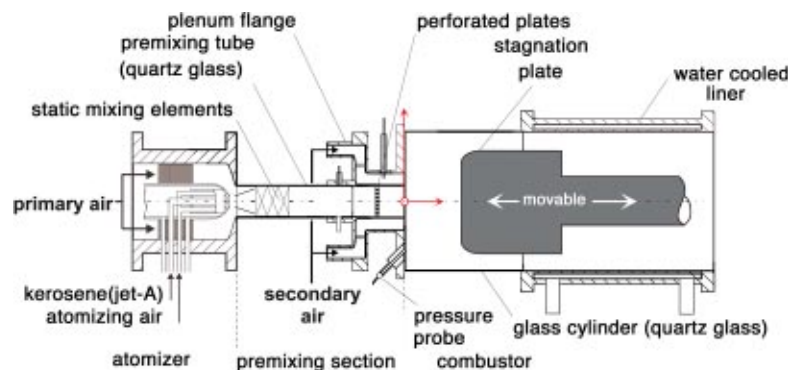


Fig. 1 Schematic of the test rig

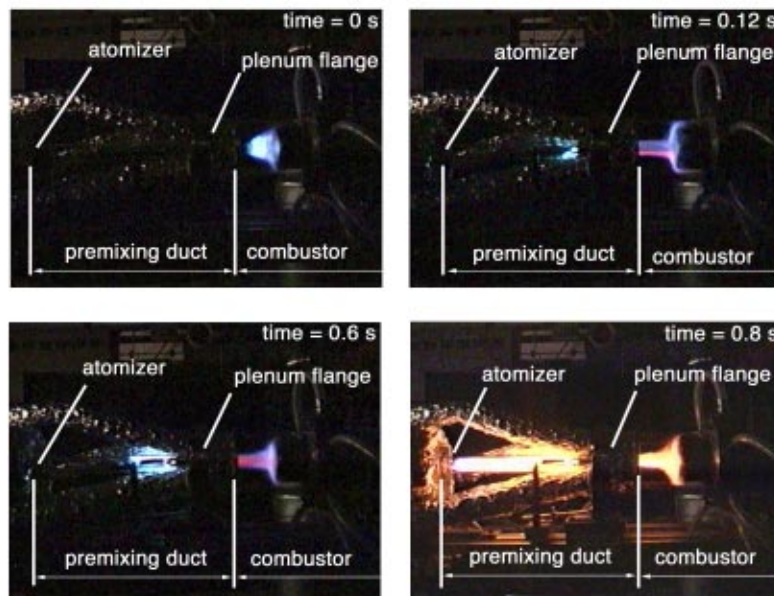


Fig. 2 Flashback from the combustor to the fuel injector

4134) in connection with a calibrated, water cooled, helium purged tube assembled to the plenum of the combustor at $x = 0$ m. Purging with helium shifts the resonance frequency of the tube to higher values and permits a correct calibration, as contamination by hot flue gases, which are a substantial problem in conventional probe measurements, is avoided. The probe permits measurements up to a sound pressure level of 160 dB. Combustion oscillations in this configuration are avoided by a suitable choice of the plate position and the ratio between the secondary and primary mass flows, [20]. The distance of the stagnation plate from the combustor inlet is within 40 to 50 mm for all the experiments presented herein.

Temperatures of the primary and the secondary air are measured with Type K thermocouples at 0.130 and 0.05 m upstream of the combustor inlet, respectively. Temperatures for the primary flow therefore include evaporation and represent the temperature of the fuel air mixture at the combustor inlet.

Measurement of Flashback

General Features. When the primary air mass flow is decreased the lean planar turbulent flame lifts off the stagnation plate, due to the decreased velocity of the unburned mixture and the increase of equivalence ratio towards stoichiometric conditions. In the case of a rapid reduction of the primary air flow and no stabilizing elements within the premixing duct, the flame flashes upstream to the atomizer, as shown in Fig. 2. During upstream propagation the flame changes from a blue premixed flame, through a state of partially premixedness with pink and orange combustion products to a highly radiative diffusion flame stabilized close to the atomizer.

When the critical conditions for flashback are approached slowly, the flame lifts off the stagnation plate and stabilizes either at the chamfered rim of the premixing tube or flashes partly back into the tube directly.

Induction of Flashback. In a first step the inlet air temperatures and average velocities are adjusted to the required values to achieve a stable flame inside the combustor. After ignition of the mixture some time is permitted to heat up the combustor. Conditions close to flashback are either achieved by decreasing the velocity of the mixture and thereby changing the equivalence ratio at the same time or by increasing the equivalence ratio towards stoichiometric conditions. The temperature at the inlet to the com-

burntor is then readjusted to account for temperature changes due to changes in the air fuel ratio. Flashback is then induced by slightly decreasing the velocity of the mixture.

Detection of Flashback. The detection of flashback was first accomplished with a photodiode triggering a fast multichannel data acquisition system. This method permitted the recording of the necessary mass flows and temperatures during flashback. As the parameters do not change during flashback, when flashback is approached slowly, flashback can be detected visually. The manual triggering of the slow acquisition system enables the recording of all the measured data.

Experimental Results

Influence of Secondary Air. For all experiments on flashback in the undisturbed pipe flow (no plates inserted) the flame propagates through the free shear layer between secondary and primary flow on the primary side (cp. Fig. 3).

As aforementioned, a slow approach of the critical conditions for flashback causes either a stabilization of the flame at the rim of the premixing tube or an upstream propagation of the flame into the tube through the wall boundary layer.

When the flame stabilizes at the rim a further reduction of the velocity leads to a complete flashback into the tube. Whether the flame stabilizes at the rim or flashes back into the tube directly depends on the conditions in the shear layer between primary and secondary air and on the flow field between the inlet and the stagnation plate, which is strongly coupled with the secondary air flow.

For equal velocities in the primary and secondary flow a planar flame is obtained. Increasing the secondary flow changes the pressure distribution in front of the stagnation plate and causes a concave shape of the flame, with the edges pointing upstream. Flame propagation into the tube is found to be independent of the ratio between primary and secondary flow, which is shown in Fig. 4. In the case of no secondary air flow ($m_s/m_p = 0$) the flame is stabilized in the shear layer close to the rim of the premixing tube by the hot exhaust gas recirculating in the corners of the combustor. In this case the data exhibit larger scattering compared to the case with secondary air. Therefore it is worthwhile to run the experiments with secondary air.

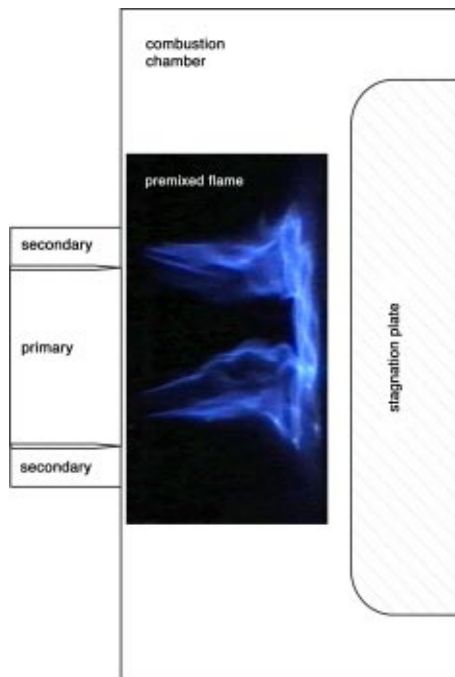


Fig. 3 Upstream flame propagation

In all cases the flame propagates upstream into the tube through the low velocity region of the wall boundary layer. The S-shaped curve in Fig. 4 exhibits a similarity to the dependence of the laminar flame velocity on the equivalence ratio. The maximum critical velocities are found for conditions close to stoichiometric, which is expected.

Influence of Homogeneity. In addition to the effects of the secondary air, the influence of the fuel distribution is investigated. The fuel distribution, in other words the local equivalence ratio, is expected to affect the upstream flame propagation into the premixing section. The influence of the homogeneity of the mixture is investigated for three configurations. In the first configuration the fuel is injected into the premixing tube with a full spray cone angle of 10 deg. In this case the fuel distribution is expected to peak on the central axis. The second configuration includes the insertion of static mixing elements into the premixing tube at about 0.1 m downstream of the 10 deg injector to achieve a homogeneous air fuel mixture at the combustor inlet. In the third

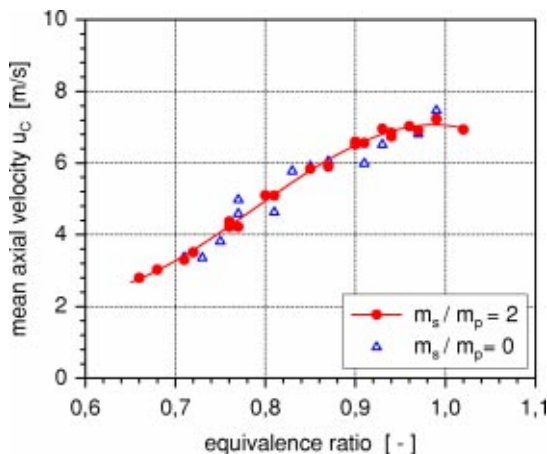


Fig. 4 Flashback into premixing duct ($T_p = T_s = 573$ K, 10 deg atomizer)

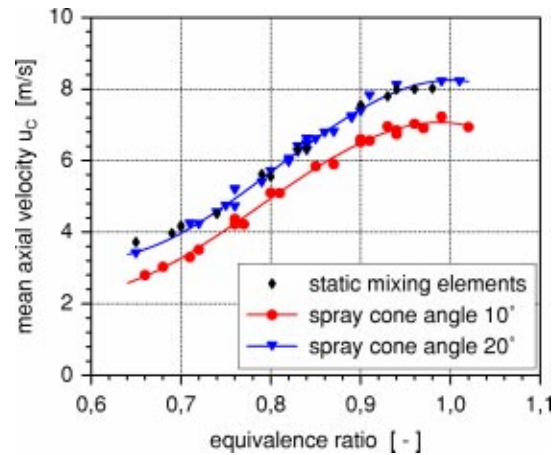


Fig. 5 Critical mean velocities at flashback for different fuel distributions ($T_p = T_s = 573$ K)

configuration an atomizer with a spray cone angle of 20 deg is applied, without mixing elements in the premixing duct.

The critical mean velocities for the three configurations are presented in Fig. 5. In the case of the 10 deg injector the critical mean velocities are lower than in the perfectly premixed case. This is due to the lean conditions close to the wall when compared to the perfectly premixed case at the same overall equivalence ratio. The highest values for the critical velocity below which flashback occurs can be detected for the perfectly premixed case. This is in accordance with the theory on flame propagation.

Influence of Temperature. Figure 6 shows the critical mean velocities for different combustor inlet mixture temperatures. Figure 6 shows that u_c appreciably increases with temperature. The maximum velocity is again found at conditions close to stoichiometric. The shape of the curves is maintained for the different temperatures. Kithrin et al. [7] proposes the flashback velocity to be approximately proportional to S_L for the case of preheating, due to the fact that the laminar flame speed S_L and the thermal diffusivity a change with temperature simultaneously. Theoretical considerations based on a constant quenching Pe number underpin this approximation. The temperature dependence of the laminar flame speed is usually given as

$$S_L \propto T^n. \quad (1)$$

Leuckel et al. [21] proposed $n = 1.93$ based on experimental data for kerosene JP-4. Assuming a constant critical Pe number (cp.

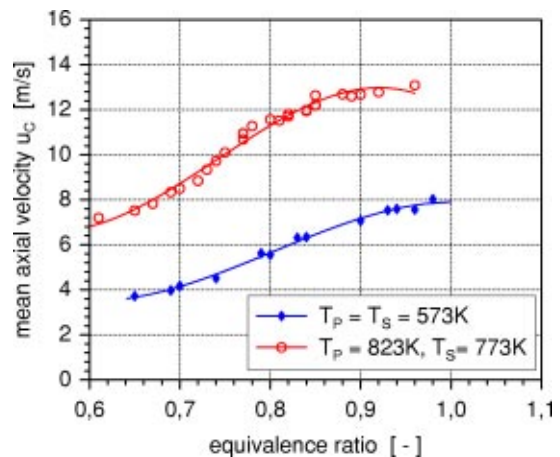


Fig. 6 Temperature dependence of flashback

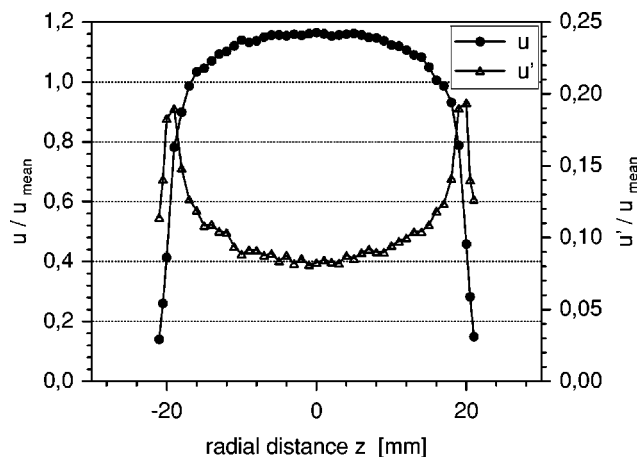


Fig. 7 Undisturbed pipe flow

p.10), the velocity gradient at the wall being proportional to the mean velocity, $S_L \propto T^{1.93}$ and $a \propto T^{1.6}$, results in a critical velocity for flashback proportional to $T^{2.16}$. In the experiment the critical velocity is found to be proportional to $T^{2.17}$ at lean conditions and $T^{1.75}$ for stoichiometric conditions. The decrease of the exponent towards stoichiometric conditions might be due to the fact that proportionality of the velocity gradient at the wall changes at larger Re numbers. Further study of those effects is needed to achieve more accurate results.

Influence of Velocity Distribution. In all configurations presented so far the velocity distribution at the inlet of the combustor represents an undisturbed, turbulent pipe flow. Typical profiles of the mean axial velocity u and the turbulence intensity u' are depicted in Fig. 7. The values are normalized with the mean axial velocity calculated from the measured mass flow, not by integration of the measured velocity profile. In all cases of a plain, undisturbed pipe flow the flame flashes back through the boundary layer. To investigate the influence of the velocity distribution and the turbulence parameters at the inlet of the combustor, perforated plates are inserted upstream of the outlet of the premixing section. Experiments with various perforated plates inserted at the two possible locations are conducted.

The following analysis will only focus on one experiment with the particular plate shown in Fig. 8, inserted at $x = -50$ mm. This plate is chosen due to the fact that a small change, namely the blockage of the central hole, changes the type of flashback, from flashback through the boundary layer to flashback through the main flow. The detailed analysis of the velocity profiles at the inlet of the combustor is realized for the case of a homogeneous fuel air mixture, stoichiometric conditions, temperatures of the primary and secondary air equal to 573 K and a mean velocity very close

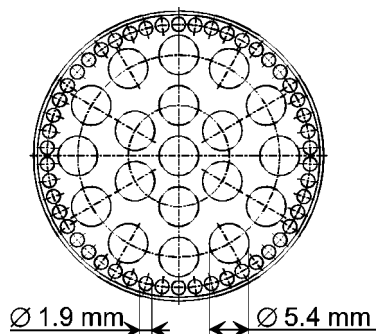


Fig. 8 Perforated plate

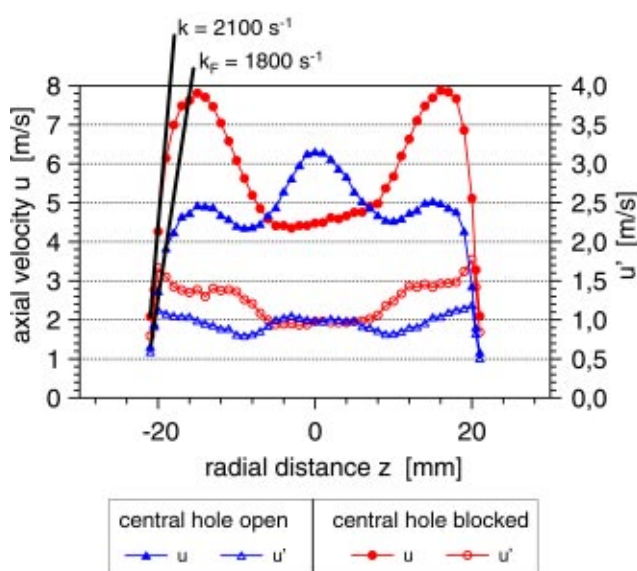


Fig. 9 Velocity distribution for flashback

to the point of flashback for both types of flashback. The velocity profiles are measured via LDV as explained earlier. The velocity distributions for both flashback situations are shown in Fig. 9. The profile of the mean velocity exhibits a maximum on the central axis and a local increase of the velocity close to the wall in the unblocked case. When the central hole is blocked the maximum on the axis vanishes and a large increase of the velocity is found close to the wall yielding a large velocity gradient at the tube wall. Around the axis the mean velocity is almost constant, with a slight increase due to the fact that the measurement line does not precisely coincide with one of the lines of symmetry of the perforated plate. The larger velocity gradient at the wall counteracts the flame propagation through the boundary layer. At the same time the reduction of the velocity on the axis permits upstream flame propagation through the main flow. Both factors lead to the transition of the flashback phenomenon.

Flashback Through the Wall Boundary Layer. In the case of an unblocked central hole, the flame propagates upstream through the wall boundary layer. In this case the critical gradient theory developed by Lewis and von Elbe (see Eq. (2) [22]) is applied to compare the measured data with theoretical considerations.

$$k_F = S_L / \delta_Q \quad (2)$$

As shown in Fig. 9, close to flashback the velocity gradient at the wall is $k_F = 1800 \text{ s}^{-1}$. The maximum value found in the experiments with the different perforated plates is $k_F = 2050 \text{ s}^{-1}$ for the homogeneous, stoichiometric flame at a preheating temperature of 573 K. For $k = 2100 \text{ s}^{-1}$ flame propagation through the wall boundary layer is inhibited (see Fig. 9). This value is used to calculate a quenching Peclet number Pe_Q (see Eq. (3)) which is compared with values in the literature.

$$Pe_Q = \delta_Q S_L \rho_u c_p / \lambda_u \quad (3)$$

The laminar flame speed S_L is 1.25 m/s, which is the maximum flame speed found via LDV measurements in a laminar kerosene stagnation flame at 573 K investigated in the same test rig. Together with the thermochemical data, Eq. (3) yields a quenching Pe number of 11.6. Values for Pe_Q for side-wall quenching of laminar flames in the literature range from 7 to 16, [23–25]. The measured value is hence in good agreement with the values found in the literature. The quenching distance and hence Pe_Q are not well-defined quantities, as they depend on the flow parameters and the boundary conditions. Curvature effects, flame stretch and ther-

mal and species diffusion influence the quenching phenomenon. For the range of Re numbers (3000–6500) necessary to induce flashback in the present geometry at atmospheric pressure, flashback through the boundary layer is predictable with the theory on laminar flashback.

Flashback Through the Main Flow. Flashback in the non-swirling flow can be treated as a competition between flashback through the boundary layer and turbulent flame propagation through the main flow. When the central hole of the perforated plate is blocked, the profile of the mean velocity shows maximum values close to the wall and an almost flat region of reduced velocity around the axis. Due to the flat profile in the vicinity of the axis an attempt to calculate the critical conditions for flashback with correlations for the flame propagation of a planar turbulent flame is self-evident. A correlation for the turbulent flame speed based on experiments of a turbulent stagnation flame is given by Eq. 4, [26].

$$S_T = S_L \left[1 + 0.435 \cdot (\text{Re}_T)^{0.44} \left(\frac{u'}{S_L} \right)^{0.4} \right] \quad (4)$$

Besides the laminar flame speed and the turbulence intensity, the turbulent Reynolds number Re_T is required in Eq. (4). Re_T is deduced from the LDV data on the central axis. Autocorrelation of the data yields the macro time scale. With Taylor's hypothesis of "frozen vortices" the macro length scale L_T is calculated to 3 mm, which is typical for this perforated plate at about 10 diameters downstream. With $\text{Re}_T=65$, $u'=1$ m/s and $S_L=1.25$ m/s Eq. (4) yields a turbulent flame speed $S_T=4.5$ m/s, which is in good agreement with the values of the mean velocity ($4.4 \text{ m/s} < u < 4.7 \text{ m/s}$) around the axis, as plotted in Fig. 9. In the present case the flame propagates upstream and stabilizes in the wakes of the perforated plate due to the fact that the turbulence intensity increases in the upstream direction, which enhances the flame speed. This analysis clearly shows that turbulent flashback through the main flow can be predicted if an appropriate correlation for the turbulent flame speed is available.

Conclusions

In the first part of generic experiments on the flashback phenomenon in LPP combustion focussing on the classical flashback, free of combustion oscillations and auto-ignition, upstream propagation of a kerosene flame has been found to strongly depend on the equivalence ratio, with maximum values for the critical velocities close to stoichiometric conditions. Fuel distribution is shown to strongly influence flashback. The maximum values for the critical velocity are found for the perfectly premixed case. Data for different preheating temperatures are shown to appreciably increase the critical velocity for flashback at higher temperatures ($T_p=823$ K). In the undisturbed, nonswirling pipe flow, the flame always propagates upstream through the low velocity region close to the wall over the whole range of Re numbers accessible in the present configuration.

Analysis of the influence of the velocity profiles at the inlet to the combustor on upstream flame propagation shows that flashback can be treated as a competition between upstream propagation through the wall boundary layer and turbulent flame propagation through the main flow. For flashback through the boundary layer the gradient theory is found to be appropriate to predict upstream flame propagation. Therefore information on the critical Pe_Q number is necessary, which depends on the flow situation, the thermal boundary conditions and the mixture properties. In the present study a critical Pe number of 15 is found to be sufficient to avoid flashback through the boundary layer. If the velocity gradient at the wall is large enough to avoid flashback through the boundary layer the flame propagates through the main flow when the turbulent flame speed exceeds the velocity in the main flow. In this case flashback is shown to be predictable with appropriate correlations for the turbulent burning velocity.

The critical mean velocities determined in the present study are much smaller than flashback data reported in the literature for practical applications, [13,17,19]. The data herein can be estimated as a limit below which upstream flame propagation is possible from stationary and stable conditions. In practical applications, combustion oscillations and complex flows, like swirling flows, have to be considered. Further studies will be needed to address those issues.

Acknowledgment

This research was supported by the Brite/Euram Low Emission Combustor Technology Program–Phase III, Aeronautics Contract BRPR-CT95-0122.

Nomenclature

Pe	=	Peclet number
Re	=	Reynolds number
S	=	flame speed (m/s)
SMD	=	Sauter mean diameter (m)
T	=	temperature (K)
a	=	thermal diffusivity (m ² /s)
c_p	=	specific heat (J/kg s)
d	=	diameter (m)
k	=	velocity gradient (l/s)
m	=	mass flow rate (kg/s)
u	=	axial velocity (m/s)
u'	=	turbulence intensity (m/s)
x	=	axial coordinate (m)
y, z	=	radial coordinate (m)

Greek Symbols

δ	=	distance (m)
λ	=	thermal conductivity (W/kg K)
ρ	=	density (kg/m ³)
ϕ	=	equivalence ratio

Indices

C	=	combustor, critical
F	=	flashback
L	=	laminar
P	=	plate, primary
Q	=	quenching
S	=	secondary
T	=	turbulent
mean	=	time mean value
u	=	unburnt

References

- [1] Tacina, R. R., 1990, "Low NO_x Potential of Gas Turbine Engines," Technical Report AIAA Paper 90-0550.
- [2] Stolarski, R. S., Baughcum, S. L., Brune, W. H., Douglass, A. R., Fahey, D. W., Friedl, R. R., Liu, S. C., Plumb, R. A., Poole, L. R., and Wesoky, H. L., 1995, "The 1995 Scientific Assessment of the Atmospheric Effects of Stratospheric Aircraft," Technical Report NASA RP 1381.
- [3] Lefebvre, A. H., 1983, "Gas Turbine Combustion," *Energy, Combustion, and Environment*, McGraw-Hill, New York.
- [4] Harris, M. E., Grumer, J., von Elbe, G., and Lewis, B., 1949, "Burning Velocities, Quenching and Stability Data on Nonturbulent Flames of Methane and Propane With Oxygen and Nitrogen," *Third Symposium on Combustion, Flame and Explosion Phenomena*, The Combustion Institute, Pittsburgh, PA, pp. 80–88.
- [5] Putnam, A. A., and Jensen, R. A., 1949, "Application of Dimensionless Numbers to Flash-Back and Other Combustion Phenomena," *Third Symposium on Combustion, Flame and Explosion Phenomena*, The Combustion Institute, Pittsburgh, PA, pp. 89–98.
- [6] Wohl, K., Kapp, N. M., and Gazley, C., 1949, "The Stability of Open Flames," *Third Symposium on Combustion, Flame and Explosion Phenomena*, The Combustion Institute, Pittsburgh, PA, pp. 3–21.
- [7] Khitrin, L. N., Moin, P. B., Smirnov, D. B., and Shevchuk, V. U., 1965, "Peculiarities of Laminar- and Turbulent-Flame Flashback," *Tenth Symposium (International) on Combustion*, The Combustion Institute, Pittsburgh, PA, pp. 1285–1291.
- [8] Bollinger, L. E., and Edse, R., 1956, "Effect of Burner-Tip Temperature on

- Flash Back of Turbulent Hydrogen-Oxygen Flames," *Ind. Eng. Chem.*, **48**(4), pp. 802–807.
- [9] Grumer, J., and Harris, M. E., 1954, "Temperature Dependence of Stability Limits of Burner Flames," *Ind. Eng. Chem.*, **46**(11), pp. 2424–2430.
- [10] Grumer, J., 1958, "Flashback and Blowoff Limits of Unpiloted Turbulent Flames," *Jet Propul.*, **28**, pp. 756–759.
- [11] Lewis, B., and von Elbe, G., 1943, "Stability and Structure of Burner Flames," *J. Chem. Phys.*, **11**, pp. 75–97.
- [12] Plee, S. L., and Mellor, A. M., 1978, "Review of Flashback Reported in Prevaporizing/Premixing Combustors," *Combust. Flame*, **32**, pp. 193–203.
- [13] Coats, C. M., 1980, "Comment on Review of Flashback Reported in Prevaporizing/Premixing Combustors," *Combust. Flame*, **37**(3), pp. 331–333.
- [14] Ghoniem, A. F., 1986, "Effect of Large Scale Structures on Turbulent Flame Propagation," *Combust. Flame*, **64**, pp. 321–336.
- [15] Keller, J. O., Vaneveld, L., Korschelt, D., Hubbard, G. L., Ghoniem, A. F., and Oppenheim, A. K., 1982, "Mechanism of Instabilities in Turbulent Combustion Leading to Flashback," *AIAA J.*, **20**(2), pp. 254–262.
- [16] Vaneveld, L., Hom, K., and Oppenheim, A. K., 1984, "Secondary Effects in Combustion Instabilities Leading to Flashback," *AIAA J.*, **22**, pp. 81–82.
- [17] Marek, C. J., Papathakos, L. C., and Verbulecz, P. W., 1977, "Preliminary Studies of Auto-ignition and Flashback in a Premixing-Prevaporizing Flame Tube Using Jet-A Fuel at Lean Equivalence Ratios," Technical Report NASA TM X-3526, Lewis Research Center, NASA, Cleveland, OH.
- [18] Proctor, M. P., T'ien, J. S., and Anderson, D. N., 1985, "Flame Flashback in a Premixed Dump Combustor," Technical Report AIAA-85-0145.
- [19] Ekstedt, E. E., and Fear, J. S., 1987, "NASA/GE Advanced Low Emissions Combustor Program," Technical Report AIAA-87-2035.
- [20] Schäfer, O., Koch, R., and Wittig, S., 2000, "Measurement of the Periodic Flow of an Enclosed Lean Premixed Prevaporized Stagnation Flame," *10th Int. Symp. on Application of Laser Techniques to Fluid Mechanics*, July, 10–13, Lisbon.
- [21] Leuckel, W., Lauer, G., Hirsch, C., and Habisreuther, P., 1994, "Mathematische Modellierung der Wechselwirkung Von Turbulenz und Reaktion unter Den in Gasturbinenbrennkammern Vorliegenden Bedingungen," Technical Report BMFT, Teilverbundprojekt Turboflam, Vorhaben 3.1.3.4.
- [22] Lewis, B., and von Elbe, G., 1987, *Combustion, Flames and Explosion of Gases*, 3rd Ed., Academic Press, San Diego, CA.
- [23] Poinso, T. J., Haworth, D. C., and Bruneaux, G., 1993, "Direct Simulation and Modeling of Flame-Wall Interaction for Premixed Turbulent Combustion," *Combust. Flame*, **95**, pp. 118–132.
- [24] Lee, S. T., and T'ien, J. S., 1982, "A Numerical Analysis of Flame Flashback in a Premixed Laminar System," *Combust. Flame*, **48**, pp. 273–285.
- [25] Lewis, B., and von Elbe, G., 1961, *Combustion, Flames, and Explosions of Gases*, 2nd Ed., Academic Press, San Diego, CA.
- [26] Liu, Y., 1991, "Untersuchung zur Stationären Ausbreitung turbulenter Vormischflammen," Ph.D. thesis, Universität Karlsruhe, Fak. f. Chemieingenieurwesen.

Time Domain Simulation of Combustion Instabilities in Annular Combustors

C. Pankiewitz¹

e-mail: pankiewitz@td.mw.tum.de

T. Sattelmayer

e-mail: sattelmayer@td.mw.tum.de

Lehrstuhl für Thermodynamik,
Technische Universität München,
85747 Garching, Germany

A novel method for the simulation of combustion instabilities in annular combustors is presented. It is based on the idea to solve the equations governing the acoustics in the time domain and couple them to a model for the heat release in the flames. The linear wave equation describing the temporal and spatial evolution of the pressure fluctuations is implemented in a finite element code. Providing high flexibility, this code in principle allows both the computational domain to be of arbitrary shape and the mean flow to be included. This yields applicability to realistic technical combustors. The fluctuating heat release acting as a volume source appears as a source term in the equation to be solved. Employing a time-lag model, the heat release rate at each individual burner is related to the velocity in the corresponding burner at an earlier time. As saturation also is considered, a nonlinearity is introduced into the system. Starting the simulation from a random initial perturbation with suitable values for the parameters of the heat release model, a self-excited instability is induced, leading to a finite-amplitude limit cycle oscillation. The feasibility of the approach is demonstrated with three-dimensional simulations of a simple model annular combustor. The effect of the model parameters and of axial mean flow on the stability and the shape of the excited modes is shown. [DOI: 10.1115/1.1582496]

Introduction

The calculation and modeling of combustion instabilities has a history reaching back to the mid of the last century, when this problem became important for the development of rocket engines, [1]. A large variety of methods has been developed since, and certainly there are different ways to classify these techniques.

Many approaches are based on the idea to describe the oscillations as a superposition of single time-harmonic modes. This allows to work in the frequency domain. In the linear case the different modes do not interact, and techniques well known from linear control theory can be applied to determine the system stability. Codes have been developed in which a complex system is split into a network of essentially one-dimensional acoustic elements, each of them transporting the acoustic pressure and velocity and, if required, additional quantities like the equivalence ratio and entropy, [2].

The Galerkin method has been applied especially in the field of rocket engines. It allows the inclusion of nonlinearities, i.e., modal coupling, [3,4]. Nonlinear acoustics are dominating here, a phenomenon usually of secondary importance for gas turbine combustors as the pressure amplitudes here are much smaller. Peracchio and Proscia [5] have adopted the approach of Culick to incorporate a nonlinear heat release model combined with linear acoustics. Akamatsu and Dowling [6] use an analytical approach to describe modal coupling for high-frequency three-dimensional thermoacoustic instabilities.

Only recently, attempts can be observed to calculate thermoacoustic oscillations in the time domain. With increasing computer performance computational fluid dynamics have proven to be applicable in principle, [7–9]. However, computational requirements are still high, and these techniques so far can only be applied to simple geometries. Wake et al. [10] have solved the one-dimensional linearized Euler equations coupled with a heat release

model, but as they do not include non-linearities the amplitudes of the unstable modes increase unboundedly. Dowling [11] has shown how for a one-dimensional geometry a nonlinear heat release model can be combined with an analytical description of the acoustics so that self-excited limit-cycle oscillations are obtained.

Though obviously many methods have been developed, so far only little progress has been made in modeling combustion instabilities in annular combustors. Walz et al. [12,13] investigate the eigenmodes of an annular combustion chamber, not considering the interaction with the flames. Krüger et al. [14,15] try to extend the one-dimensional network technique to annular combustors, while Krebs et al. [16] show how the Galerkin method can be applied in principle. Stow and Dowling [17] and Evesque and Polifke [18] employ low-order modeling techniques. Nevertheless the coupling of more than one heat release zone with multidimensional acoustics still is not clearly understood.

It is therefore appealing to simulate combustion instabilities in an annular combustor in the time domain. If a description of the heat release in terms of the acoustic quantities is available, no further assumptions about the modal coupling have to be made. We have developed such an approach, combining the linear wave equation describing the acoustics with a nonlinear heat release model. The following section introduces the governing equations and the numerical tool to solve them. We then present a model annular combustor used for exemplary simulations to verify the feasibility of the technique. The results of these simulations are given in the subsequent section. In the conclusions we also shortly discuss how this new method will be extended in future.

Description of the Technique

Governing Equations. Combustion instabilities in gas turbine combustors are known to exhibit pressure and velocity fluctuations whose amplitudes even in the case of strong oscillations are small compared to the mean pressure and velocity of sound, respectively. As additionally the attenuation of acoustic waves due to the viscosity of the fluid can be neglected, the linearized Euler

¹To whom all correspondence should be addressed.

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Amsterdam, The Netherlands, June 3–6, 2002; Paper No. 2002-GT-30063. Manuscript received by IGTI, Dec. 2001, final revision, Mar. 2002. Associate Editor: E. Benvenuti.

equations provide a suitable description of the acoustics in the combustor. These equations are derived from the conservation of mass²,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0, \quad (1)$$

and momentum,

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} = -\frac{1}{\rho} \nabla p, \quad (2)$$

when the pressure $p = \bar{p} + p'$, density $\rho = \bar{\rho} + \rho'$ and velocity $\mathbf{u} = \bar{\mathbf{u}} + \mathbf{u}'$ each are written as the sum of their mean and their fluctuating part. The fluctuations of density can be attributed to the fluctuations of pressure and entropy. If the entropy fluctuations are solely caused by heat release fluctuations, the following equation holds for the substantial derivative of the density with respect to time (cf. [19]):

$$\frac{D\rho}{Dt} = \frac{1}{a^2} \frac{Dp}{Dt} - \frac{\kappa - 1}{a^2} \dot{q}. \quad (3)$$

Here, $a = \sqrt{\kappa p / \rho}$ is the velocity of sound, κ is the ratio of specific heats, and $\dot{q} = \bar{\dot{q}} + \dot{q}'$ is the rate of heat addition per unit volume. Combining Eq. (3) with the continuity equation Eq. (1), linearizing the resulting expressions, and using the relation for isentropic changes of state $p' / \rho' = a^2$, yields the linearized equations for the conservation of mass,

$$\frac{\partial p'}{\partial t} + \mathbf{u}' \cdot \nabla \bar{p} + \bar{\mathbf{u}} \cdot \nabla p' + \kappa p' \nabla \cdot \bar{\mathbf{u}} + \kappa \bar{p} \nabla \cdot \mathbf{u}' = (\kappa - 1) \dot{q}', \quad (4)$$

and momentum,

$$\frac{\partial \mathbf{u}'}{\partial t} + (\bar{\mathbf{u}} \cdot \nabla) \mathbf{u}' + (\mathbf{u}' \cdot \nabla) \bar{\mathbf{u}} + \frac{p'}{\rho a^2} (\bar{\mathbf{u}} \cdot \nabla) \bar{\mathbf{u}} = -\frac{1}{\rho} \nabla p'. \quad (5)$$

These equations do not account for entropy waves, which can lead to some inaccuracy only for very low frequencies. For a discussion of this topic see Sattelmayer [2] and Polifke et al. [20].

Equations (4) and (5) can further be simplified for the case of uniform mean flow. Omitting the terms involving the spatial derivatives of the mean quantities, applying $\bar{\rho} a^2 \nabla \cdot$ to the momentum equations Eq. (5) and subtracting the result from the substantial derivative of the continuity equation Eq. (4) with respect to time leads to the wave equation

$$\begin{aligned} \frac{1}{a^2} \left(\frac{\partial^2 p'}{\partial t^2} + 2 \bar{\mathbf{u}} \cdot \nabla \left(\frac{\partial p'}{\partial t} \right) + \bar{\mathbf{u}} \cdot (\bar{\mathbf{u}} \cdot \nabla) (\nabla p') \right) - \bar{\rho} \nabla \cdot \left(\frac{1}{\rho} \nabla p' \right) \\ = \frac{\kappa - 1}{a^2} \frac{D \dot{q}'}{Dt}. \end{aligned} \quad (6)$$

Compared to Eqs. (4) and (5) this equation has the advantage that we only have to solve for the acoustic pressure p' but not for the three components of the sound particle velocity $\mathbf{u}'$. However, Eq. (6) disregards loss mechanisms like area jumps. This simplification and neglecting entropy waves does not mean a general restriction of the technique, though. The governing equations could be extended to include the mentioned effects. In the framework of this work we focus on the purely acoustical mechanisms. The substantial derivative of the heat release rate fluctuations in Eq. (6) implies that changes in the heat release can be caused both by local changes of the heat release rate and by a moving heat release zone. Equation (6) can be solved if a relation describing the coupling of the heat release to the flow field is provided. Giving a very general formulation of that relation, the fluctuating heat release may depend on both the pressure and the velocity, additionally allowing for a time delay:

$$\dot{q}'(\mathbf{x}, t) = \dot{q}'((p, \mathbf{u})(\mathbf{x}, t, \tau(\mathbf{x}))). \quad (7)$$

²Notation: $\nabla \cdot \mathbf{u} = u_x + v_y + w_z$, $\nabla \mathbf{u} = (u_x, v_x, w_x)^T$

Here, the delay time τ does not necessarily have to be a single value but can rather be a delay time distribution. This may be important for a realistic model of a gas turbine burner flame as has been shown by Sattelmayer [2] and Flohr et al. [21]. Equation (6) involves p' only linearly. In order to make unstable perturbations not grow into unlimited oscillations, it is therefore important to include nonlinearities occurring in the heat release.

Our technique for the simulation of combustion oscillations can be summed up as follows: provide a nonlinear model for the fluctuations of the heat release rate (Eq. (7)) to solve the wave equation Eq. (6) in the time domain together with appropriate boundary conditions. The following paragraph will give a short description of the numerical tool we use to realize this idea.

Numerical Tool. As in the case of combustion oscillations in annular combustors instabilities in the low-frequency regime are dominating, and their wave lengths are of the same order as the characteristic lengths of the combustor, even standard numerical methods can lead to satisfying results. To solve the equations described before, we employ the commercial tool FEMLAB. This is a finite element package designed to solve an arbitrary set of partial differential equations of the form

$$d_a \frac{\partial \mathbf{u}}{\partial t} - \nabla \cdot (c \nabla \mathbf{u} + \alpha \mathbf{u} - \gamma) + \beta \nabla \mathbf{u} + a \mathbf{u} = \mathbf{f} \quad \text{in } \Omega, \quad (8)$$

$$\mathbf{n} \cdot (c \nabla \mathbf{u} + \alpha \mathbf{u} - \gamma) + q \mathbf{u} = g - h^T \lambda \quad \text{on } \partial \Omega, \quad (9)$$

$$h \mathbf{u} = r \quad \text{on } \partial \Omega. \quad (10)$$

Here, $\mathbf{u}$ denotes a vector of solution components in a domain Ω . The coefficients d_a , c , α , γ , β , a , f , q , g , h , and r are scalars, vectors, matrices, or tensors and can be complex-valued functions of the space, time, the solution $\mathbf{u}$ and, except for q , g , h , and r , also of the gradient of $\mathbf{u}$, [22]. Equations (9) and (10) constitute a generalized Neumann boundary condition and a Dirichlet boundary condition, respectively, where $\mathbf{n}$ is the outward normal vector at the boundary, and λ is a Lagrange multiplier. The domain Ω is triangulated with a finite element mesh consisting of tetrahedral elements. The general formulation of the PDEs in FEMLAB, together with the possibility to model arbitrary geometries, opens the way to the solution of complex problems.

In order to match the form prescribed by Eq. (8), the wave equation Eq. (6) is rewritten as two first-order differential equations. The solution of the resulting system of equations is based on the method of lines. For the time integration of the resulting ODEs we employ an implicit solver with automatic step size control implementing a second-order backward differentiation formula (BDF) method.

Model Annular Combustor

Geometry. To verify the feasibility of our approach, we have chosen to carry out calculations in a generic geometry. This model is derived from an annular combustor test rig that in future will give us valuable experimental information for the understanding of combustion instabilities and verification of models (see Fig. 1). The test rig consists of an annular plenum chamber and an annular combustion chamber connected by 12 swirl burners. The plenum chamber is supplied with an externally premixed fuel-air mixture, so fuel inhomogeneities do not occur and entropy waves do not play a role in this combustor. The hot gas exits the combustor through 12 nozzles, allowing operation at slightly elevated pressure to establish a choked end if desired.

In the numerical model we avoid a detailed reproduction of the swirl burners, exit nozzles, and inlet orifices. The leads to a simple geometry consisting of an inlet plane, followed by an annular plenum, 12 straight ducts representing the swirler section, the annular combustion chamber, and finally an exit plane (see Fig. 2).

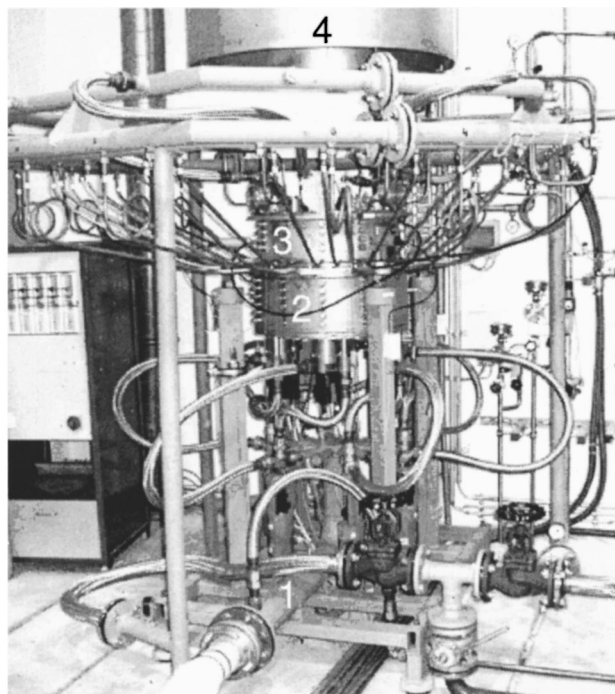


Fig. 1 Annular combustor test rig representing the basis for the model combustor. (1) Premixing section, (2) plenum chamber, (3) combustion chamber, (4) exhaust.

Boundary Conditions. At the walls, the outward normal component of the sound particle velocity $\mathbf{n} \cdot \mathbf{u}'$ has to disappear. This is ensured by the boundary condition

$$\mathbf{n} \cdot \nabla p' = 0. \quad (11)$$

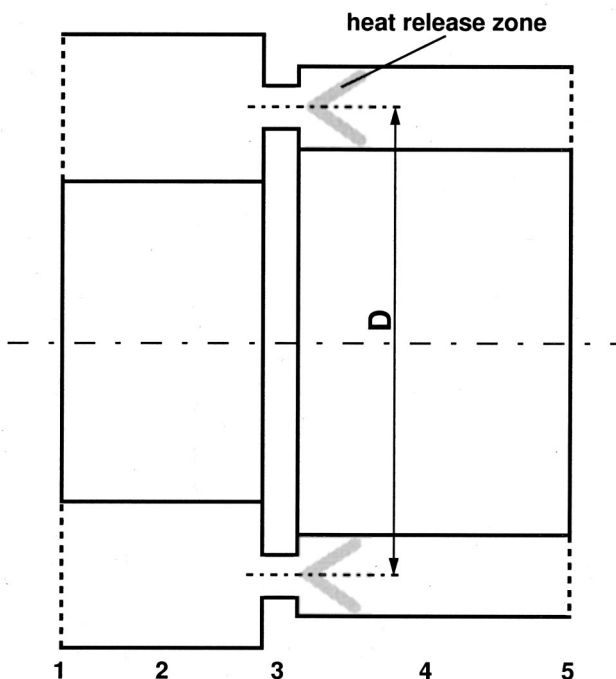


Fig. 2 Geometry of the model annular combustor (cross section) with (1) inlet, (2) plenum chamber, (3) burner section, (4) combustion chamber, and (5) exit

For the inlet and outlet boundary conditions we follow Akamatsu and Dowling [6] and assume that the flow both enters and exits the computational domain from and to a large plenum so that pressure perturbations there can be neglected. Then the outward normal component of the velocity fluctuation will be proportional to the pressure fluctuation:

$$\mathbf{n} \cdot \mathbf{u}' = K p'. \quad (12)$$

Using the momentum equation in the normal direction, this translates to

$$\mathbf{n} \cdot \nabla p' = - \frac{K}{K \bar{u}_n + \frac{1}{\rho}} \frac{\partial p'}{\partial t}, \quad (13)$$

where $\bar{u}_n$ is the normal component of the mean velocity (taken positive in the outward direction). The constant $K = u_n / 2(p - p_\infty)$ can be determined from the pressure drop across the inlet or exit, respectively. Except for $K=0$ (closed or choked end in the limiting case of zero Mach number mean flow) and $K \rightarrow \infty$ (open end), only parts of the acoustic waves are reflected at these boundaries. This establishes a loss of acoustic energy, which is important for the evolution of a limit cycle as will be shown in the results section.

Nonlinear Heat Release Model. A variety of flame and heat release models is available that could be used to provide a relation for Eq. (7). The assumptions and requirements underlying the model employed here are as follows:

1. The zone of heat release is a predefined fixed region at each burner. There is no movement of the flame.
2. The instantaneous rate of heat released per unit volume at each burner is related to the burner exit velocity u_B at the corresponding burner, involving a single delay time τ .
3. The response of the flame to velocity fluctuations is increasingly damped with increasing frequency.
4. The maximum and minimum heat release are limited, i.e., saturation effects introduce a nonlinearity.

Such a simple model of course is far from describing a realistic technical flame. However, it still is suitable for verifying our approach and investigating basic mechanisms of the combustion instabilities.

A model fulfilling the aforementioned points was used by Dowling [11]. We basically adopt it and write for the fluctuation of the heat release rate at one burner j

$$\frac{1}{\omega_0} \frac{d \dot{q}'_j}{dt} + \dot{q}'_j = \eta f(u'_{B,j}(t - \tau)), \quad j = 1, \dots, 12, \quad (14)$$

where $\eta = \bar{\dot{q}} / \bar{u}_B$ is the ratio of the mean heat release rate to the mean burner exit velocity, and the function f describes how the heat release rate fluctuates proportionally to the velocity, also considering saturation (see also Fig. 3):

$$f(u) = \begin{cases} u & \text{if } |u| \leq \sigma \bar{u}_B \\ \text{sign}(u) \sigma \bar{u}_B & \text{if } |u| > \sigma \bar{u}_B \end{cases}. \quad (15)$$

In contrast to Dowling, we introduce the additional parameter σ which allows to “turn on” saturation already for fluctuations that are smaller than the mean value. This seems meaningful as a real flame usually will be driven into the nonlinear regime even for moderate fluctuations. A very extreme example has been given by Hantschk and Vortmeyer [9]. Their numerical simulations show that the total heat release rate might undergo nonlinear modifications even for fluctuations which are only a few percent of the mean value. There, this most likely has to be attributed to the fact that the flame is very long so that regions with strong positive fluctuations could compensate regions with strong negative fluctuations. Though we do not account for a spatial variation of the

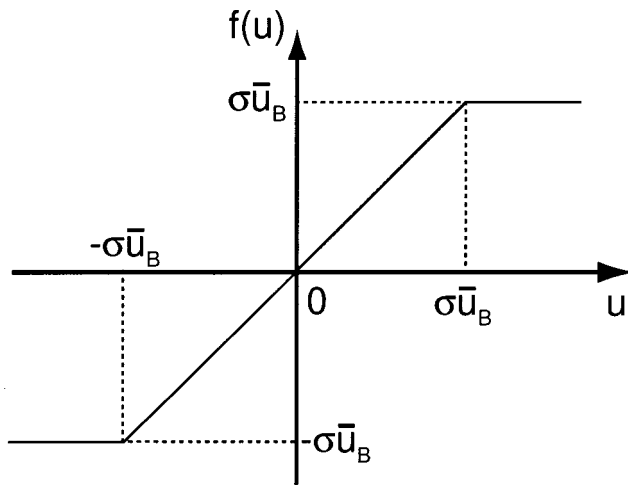


Fig. 3 Shape of the function $f(u)$ defined by Eq. (15)

heat release rate and time delay, such effects to some extent can be modeled with the parameter σ . For all calculations presented here, $\sigma=0.75$.

The damping of higher frequency oscillations also leads to a saturation amplitude decreasing with increasing frequency. This has experimentally been observed for example by Harper et al. [23]. Such an effect is also modeled with the time derivative in Eq. (14), ω_0 being the cutoff frequency of the low-pass filter.

The region of heat release is compact compared to the wave length of the instabilities expected. As additionally the delay time has a single value, we expect the actual spatial distribution of the heat release rate to be of minor importance and decide to model the swirl stabilized flames very simply with conical shapes (see Fig. 2).

To evaluate Eq. (14) the burner exit velocity fluctuation u'_B has to be known. It can easily be computed from the pressure fluctuation field p' using the continuity and momentum equations Eqs. (4) and (5) locally, provided there is either no or purely axial mean flow. For no mean flow integration of the momentum equation in axial direction at the burner exit directly yields u'_B . For axial mean flow first the two other momentum equations have to be used in the same way at adjacent grid points to determine the velocity components normal to u'_B and their gradients, in order to subsequently solve for u'_B using the axial momentum equation and the continuity equation.

At each time step carried out for the solution of Eq. (6), the differential equations Eq. (14) are integrated numerically to provide the right-hand side of Eq. (6). The ODE solver used implements a pair of embedded Runge-Kutta schemes of order four and five (Dormand-Prince pair).

Dimensionless Quantities and Mean Conditions. All results in this paper will be given in a dimensionless form. For this purpose we use the mean diameter $D=0.437$ m of the combustor as the characteristic length L_0 and the mean velocity of sound in the plenum chamber $\bar{a}_1=556$ m/s as the characteristic velocity u_0 . Additionally employing the characteristic impedance of the fluid in the plenum chamber, $Y_0=\bar{\rho}_1\bar{a}_1=250$ Pa s/m, the other characteristic quantities can be derived: time $t_0=L_0/u_0$, frequency $f_0=1/t_0$, pressure $p_0=Y_0u_0$, density $\rho_0=Y_0/u_0$ and volumetric heat release rate $\dot{q}_0=Y_0u_0^2/L_0$. All values given below are nondimensionalized with these quantities.

The mean (steady-state) conditions are assumed to be piecewise constant in the different sections of the combustor. This means the velocity of sound and the density are $\bar{a}_1$ and $\bar{\rho}_1$ throughout the plenum chamber and the burners (ducts), and $\bar{a}_2$ and $\bar{\rho}_2$ in the combustion chamber. The ratio of specific heats enters the gov-

erning equations only in the heat release zone and there is taken to be constant having the value of the burned gas-air mixture, $\kappa=1.33$.

Most of the results presented later are calculated for no mean flow conditions, i.e., $\bar{\mathbf{u}}=0$. The cases where mean flow is included are limited to axial mean flow, $\bar{u}_1$, $\bar{u}_B$, and $\bar{u}_2$ being the velocities in the plenum chamber, burners and combustion chamber, calculated from the mean mass flow rate, mean density, and the cross-sectional areas.

The mean conditions have been set according to a case where the experimental combustor would be operated with natural gas, the equivalence ratio being $\phi=0.69$, the fuel-air mixture being preheated to $T_1=770$ K, and the total thermal power being $P=1020$ kW. Then the parameters entering the model become

$$\bar{a}_1=1 \quad \bar{a}_2=1.7 \quad \bar{\rho}_1=1 \quad \bar{\rho}_2=0.33 \quad \eta=61.4$$

$$K_{in}=0.1 \quad K_{out}=0.2 \quad \bar{u}_1=0.011 \quad \bar{u}_B=0.14 \quad \bar{u}_2=0.06$$

Here K_{in} and K_{out} refer to the constant described in the boundary conditions section at the inlet and outlet, respectively.

Eigenmode Analysis. It will be useful to compare the results from the time domain simulations with the eigenmodes of the combustor. To compute these eigenmodes, we take Eq. (6) with the right-hand side set to zero and assume a solution for the pressure fluctuations $p'=\hat{p}(\mathbf{x})e^{i\omega t}$. This leads to the eigenvalue problem

$$\Delta \hat{p} - \frac{1}{\bar{a}^2} \bar{\mathbf{u}} \cdot (\bar{\mathbf{u}} \cdot \nabla) (\nabla \hat{p}) - i \frac{2}{\bar{a}^2} \omega \bar{\mathbf{u}} \cdot \nabla \hat{p} + \frac{\omega^2}{\bar{a}^2} \hat{p} = 0. \quad (16)$$

As the eigenvalue ω here appears both linearly and quadratically, it is not straightforward to solve Eq. (16) numerically with standard methods. This problem can be overcome by substituting Eq. (16) by a set of two equations,

$$\tilde{p} = \frac{\omega}{\bar{a}} \hat{p}, \quad (17)$$

$$\Delta \tilde{p} - \frac{1}{\bar{a}^2} \bar{\mathbf{u}} \cdot (\bar{\mathbf{u}} \cdot \nabla) (\nabla \tilde{p}) - i \frac{2}{\bar{a}} \bar{\mathbf{u}} \cdot \nabla \tilde{p} + \frac{\omega}{\bar{a}} \tilde{p} = 0, \quad (18)$$

which both involve ω only linearly. Then the boundary conditions Eqs. (11) and (13) become

$$\mathbf{n} \cdot \nabla \hat{p} = 0 \quad (19)$$

and

$$\mathbf{n} \cdot \nabla \tilde{p} = -i\bar{a} \frac{K}{Ku_n + \frac{1}{\rho}} \tilde{p}. \quad (20)$$

Equations (17) to (20) constitute a standard eigenvalue problem that can readily be solved within FEMLAB which also provides an eigenvalue solver.

The eigenfrequencies f and periods T of the first four eigenmodes of the model combustor are

Mode	(0,1,0)	(1,0,0)	(1,1,0)	(0,2,0)
f	0.353	0.390	0.656	0.668
T	2.83	2.56	1.53	1.50

Mode (l,m,n) in this table refers to the contribution of modes of order l , m , and n in axial, circumferential, and radial direction, respectively. The eigenfrequencies for the cases with and without mean flow are virtually identical, so just one frequency is given. This has to be attributed to the low Mach numbers which are 0.011, 0.14, and 0.035 in the plenum, burner section, and combustion chamber, respectively.

The pressure distribution for these first four modes can be seen in Fig. 4. Note that the modes involving a circumferential compo-

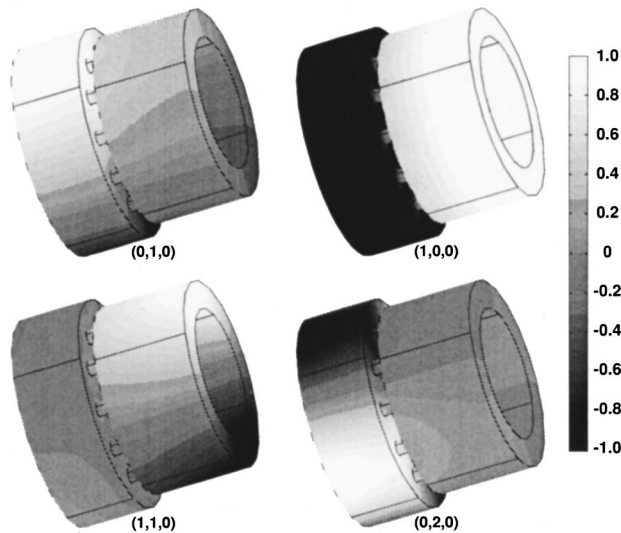


Fig. 4 First four acoustical eigenmodes of the combustor (normalized pressure distribution)

nent are doubly degenerate, i.e., there belongs a two-dimensional eigenfunction space to one eigenvalue. For reasons of simplicity we only show one of the two eigenfunctions. The second is identical to the first rotated by $\pi/(2m)$ around the axis of the combustor, m being the order of the circumferential mode.

When there is a circumferential component, for low frequencies the mode is prominent only either in the plenum chamber or in the combustion chamber, not in both. This is caused by the different velocities of sound in the two annuli which are not integer multiples of each other. Here, the first two pure circumferential modes have strong amplitudes only in the plenum chamber. What might appear as the first circumferential mode just of the combustion chamber under close examination reveals to be the first combined axial-circumferential mode, its amplitude in the plenum chamber only being very low.

FEM Mesh and Initial Condition. The combustor geometry has been discretized with a finite element mesh consisting of approximately 70,000 tetrahedral elements. This mesh is shown in Fig. 5.

In this figure also the pressure distribution used as the initial condition can be seen. The pressure fluctuation at the node points is randomly set to values between -0.05 and 0.05 . This generates an initial perturbation with a broad spectrum of wave numbers that is expected to trigger the growth of unstable modes.

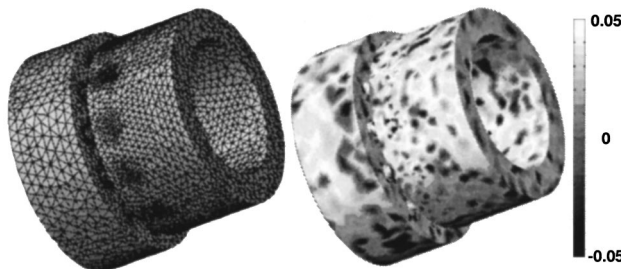


Fig. 5 FEM mesh and initial distribution of pressure fluctuation p'

Results

Limit Cycle Evolution. In the time-domain simulation, the random perturbation in the pressure field imposed as an initial condition induces a self-excited instability if the parameters of the heat release model are set appropriately. This can be seen in Fig. 6 where the fluctuation of the burner exit velocity at one specific burner in terms of the mean burner exit velocity is shown. The calculation has been carried out for no mean flow conditions with a delay time $\tau=1.9$ and a cutoff frequency $\omega_0=5$. The initially random fluctuations turn into a periodic oscillation which at first grows exponentially. When the velocity fluctuations become so strong that saturation occurs, the growth rate gets smaller until finally the amplitude remains constant. At this point the limit cycle is reached.

More information about this limit cycle can be obtained from the evolution of the pressure distribution during one period of the oscillation, shown in Fig. 7. It turns out that the mode excited in this case is the first combined axial-circumferential mode $(1,1,0)$. Here this mode appears as a spinning mode, i.e., the node line is not stationary but rather rotates by 2π during one cycle. This leads to a phase difference between the fluctuations at the burners and the combustor exit, which best can be recognized from the fact that the node line is oblique and not parallel to the combustor axis. The pressure at the exit lags the pressure at the burners because the information takes a certain time to travel from the heat release zone to the exit.

The mechanism underlying the self-excited instability becomes clearer examining the relation of the burner exit velocity fluctuations, pressure fluctuations, and heat release fluctuations at one burner. Figure 8 shows these quantities in terms of their mean values for a few periods of the limit cycle. Here, the pressure fluctuation at the center of the heat release zone of the corresponding burner is given. It can be seen that the amplitude of the heat release fluctuation is limited to a value slightly below $0.75\bar{q}$ which is the value given for the onset of saturation. This value is not fully reached because of the additional damping in the heat release model. The velocity fluctuations, however, exceed the saturation limit and gain an amplitude just above the mean velocity, meaning that there is a short part of the cycle where reflow can be observed. The pressure fluctuations reach about 10% of the mean pressure. This is a value probably too high for a real combustor, but our model so far does not include any other loss mechanisms than the sound transmission at the inlet and exit. Including other loss effects would very likely yield lower pressure fluctuations.

From Fig. 8, it is also obvious that the pressure fluctuation lags the burner exit velocity fluctuation almost exactly by $T/4$, $T=1.6$ being the period of the oscillation. This implies an almost purely imaginary impedance at the location of the flame. As in

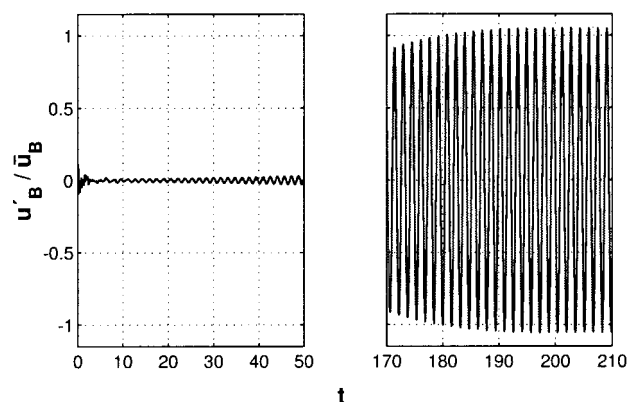


Fig. 6 Time evolution of a self-excited instability for $\tau=1.9$, $\omega_0=5$, no mean flow

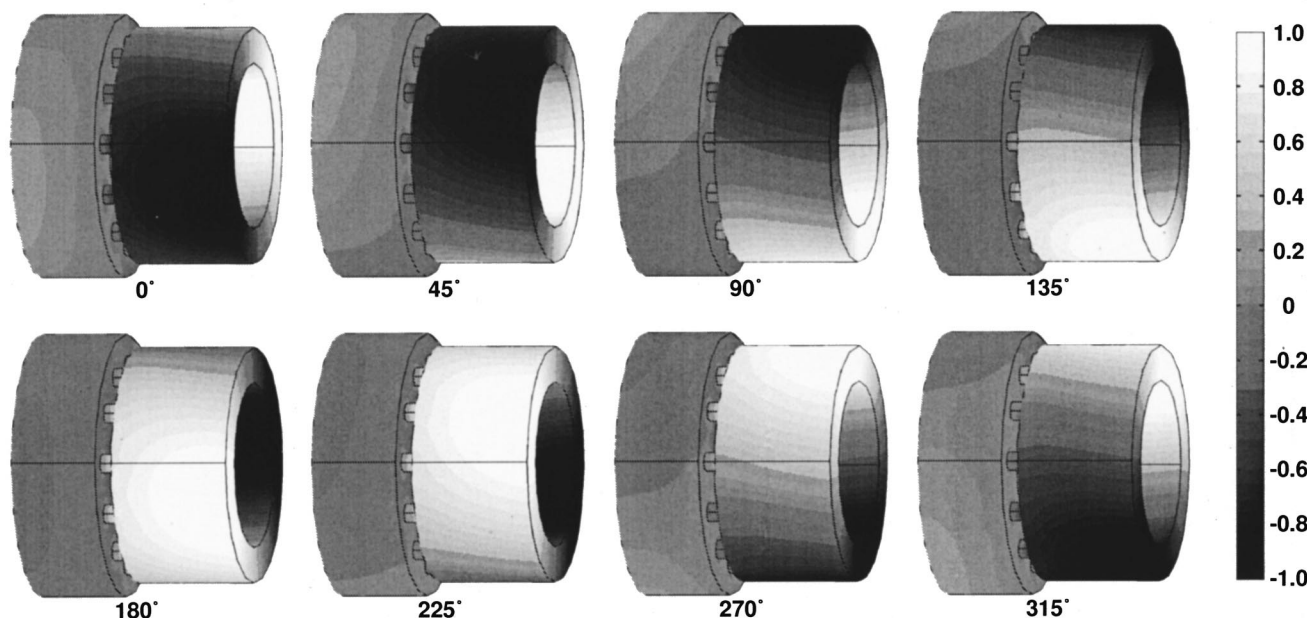


Fig. 7 Normalized pressure distribution at different phase angles for one period of the limit cycle for $\tau=1.9$, $\omega_0=5$, no mean flow

turn the heat release fluctuation lags u'_B by $\tau=1.9$, which is close to $5/4T$, it is nearly in phase with the pressure fluctuation. This clearly means that the Rayleigh criterion is fulfilled, and the oscillation is excited or, during the limit cycle, sustained.

The existence of saturation and thus a nonlinearity leads to the production of higher frequency components. This can be seen in the spectrum in Fig. 9 which was computed from the burner exit velocity fluctuations at one burner. Besides a strong component at 0.637 which is close to the frequency of the limit cycle oscillation, $f=0.625$, odd multiples of this value can be found. This is the consequence of the clipping of the heat release rate. The amount of energy transferred to the higher harmonics rapidly decreases with increasing frequency.

It is worth paying some attention to the mechanism establishing a finite-amplitude limit cycle. For this reason, Fig. 10 shows the time evolution of the burner exit velocity and heat release rate fluctuations for a case where K_{in} and K_{out} have been set to zero so that there are absolutely no losses in the system (except for numerical dissipation). It turns out that, though the heat release rate clearly is bounded, the acoustic quantities grow unboundedly but linearly. If the Rayleigh criterion is fulfilled, a limited heat release

fluctuation means that during each cycle of the oscillation a constant amount of energy is added to the oscillation, which leads to a linear growth of the velocity and pressure fluctuations. Only when losses are existent, which grow quadratically with the acoustic quantities, a limit cycle can be established when the energy gain in the flames equals the energy losses. It therefore must be concluded that a time-domain simulation aiming at reproducing limit cycle oscillations carefully has to consider the loss mechanisms. Theoretically, a heat release model which in the saturation regime would produce decreasing heat release rate fluctuations with increasing velocity fluctuations would also result in a finite-amplitude limit cycle, [24], but it is at least questionable if such a model has any relevance to a realistic flame.

The cases shown so far exhibit a self-excited instability. If the model parameters are chosen appropriately, also the stable case can be observed as shown in Fig. 11 for $\tau=1.5$ and $\omega_0=5$. The initial perturbations decay exponentially and ultimately vanish. Nevertheless, the initially random fluctuations still undergo a transition to an ordered oscillation so that with increasing time a "least stable" mode becomes prominent.

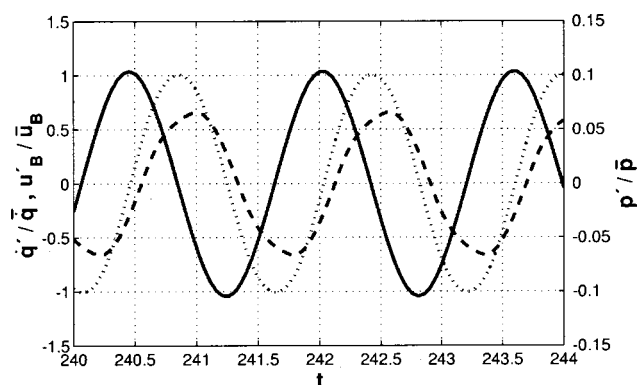


Fig. 8 Fluctuation of burner exit velocity (—), pressure (····), and heat release rate (— — —) during the limit cycle for $\tau=1.9$, $\omega_0=5$, no mean flow

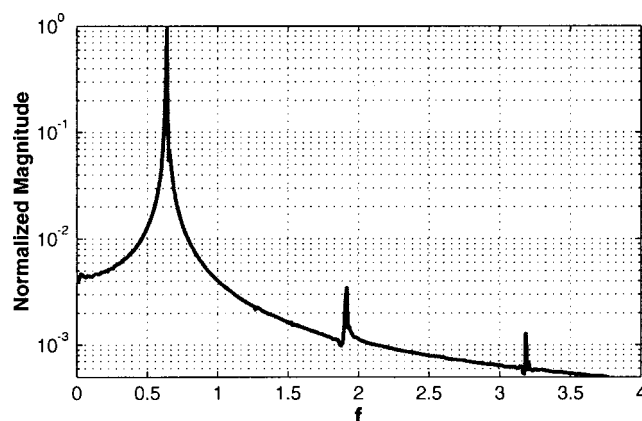


Fig. 9 Normalized frequency spectrum of the limit cycle oscillation for $\tau=1.9$, $\omega_0=5$, no mean flow

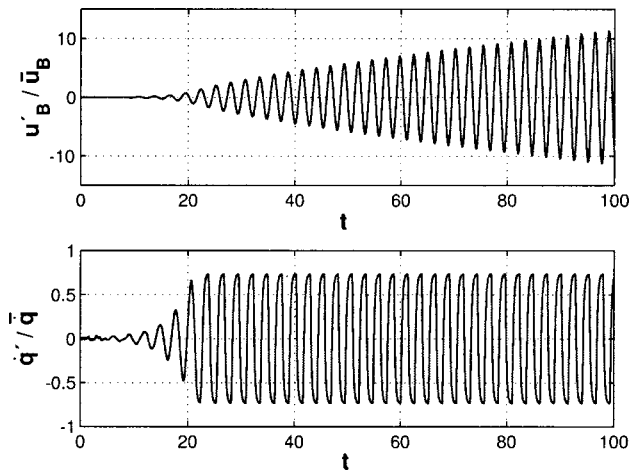


Fig. 10 Time evolution of a self-excited instability in a system without losses ($\tau=1.0$, $\omega_0=5$, no mean flow)

Stability and Mode Types. In order to investigate the influence especially of the heat release model on the stability and mode types of the combustor, we have carried out calculations for different delay times in the range from $\tau=1.0$ to $\tau=2.5$ for a fixed cutoff frequency $\omega_0=5$. The stability can be assessed looking at the cycle increment. This is a measure for the part of the amplitude by which the oscillation grows or decays during one cycle. Positive values correspond to a self-excited instability, negative values to a damped mode. The cycle increment obtained from our calculations can be found in Fig. 12. Clearly, the stability strongly depends on the value of the delay time. For $\tau=1.9$ we also present two cases with a decreased cutoff frequency, $\omega_0=4$ and $\omega_0=3$, respectively. Obviously the stronger damping has a stabilizing effect so that the unstable mode can be turned into a stable mode.

The effect of the delay time becomes even more obvious when we examine the frequency and mode type corresponding to each case. Figure 13 shows the frequency of the oscillation as a function of the delay time for $\omega_0=5$. Additionally the type of the (unstable and least stable, resp.) mode is given. Either the pure axial mode (for the corresponding pressure distribution see Fig. 4) or the first combined axial-circumferential mode already described before is observed in distinct regions for the value of τ . At the borders of these regions the corresponding modes are stable while they become unstable further away. This means that a zone of stability separates regimes of different modes that grow into a self-excited limit cycle instability.

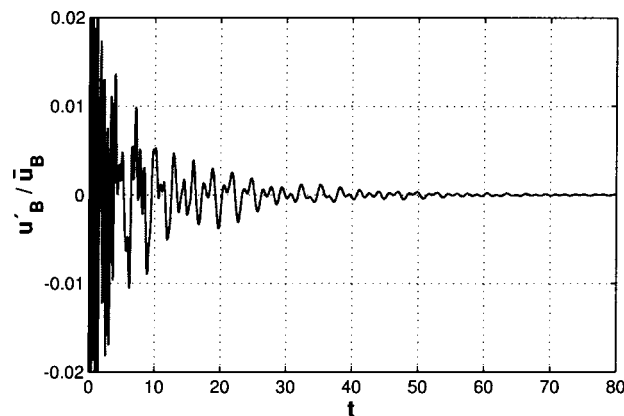


Fig. 11 Time evolution of the velocity fluctuation at one burner for $\tau=1.5$, $\omega_0=5$, no mean flow

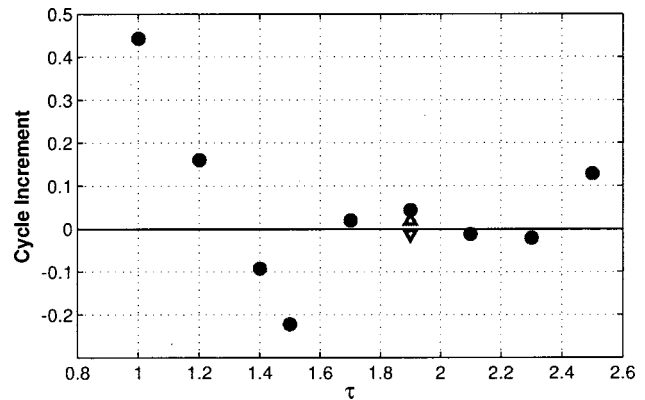


Fig. 12 Cycle increments for different delay times and $\omega_0=5$ (●), $\omega_0=4$ (△), $\omega_0=3$ (▽), no mean flow

In Fig. 13 for the two modes observed also the eigenfrequencies of the acoustical eigenmodes (i.e., without coupling to the heat release zone) calculated above are given. Though the frequencies of the combustion oscillations lie somewhere in the vicinity of these eigenfrequencies, the deviations generally are considerable and cannot be neglected. A pure eigenmode analysis of an annular combustor can therefore only give very limited information regarding the problem of combustion instabilities.

The results presented so far all refer to no mean flow conditions. The eigenmode analysis already has indicated that due to the low Mach numbers of the mean flow its influence should be of minor importance. To verify this assumption, we have carried out two calculations for $\tau=1.0$ and $\tau=1.9$, respectively, and $\omega_0=5$, where the mean flow has been included. Indeed the frequencies and cycle increments obtained are virtually identical to the no mean flow case. This must be attributed to the fact that the low Mach number effects to the acoustics can be neglected. The situation would probably be different if we included mean flow-dependent loss mechanisms and entropy waves.

However, there still is an interesting difference. While for the case with $\tau=1.9$ and $\omega_0=5$ the mode obtained for no mean flow is a spinning mode as already discussed, in the case with mean flow it appears as a standing mode with a stationary, axis-parallel node line but identical frequency, see Fig. 14. We have observed the standing mode in two independent calculations with different random initial conditions and mean flow included, while all of the no mean flow cases have shown the spinning mode. Additionally,

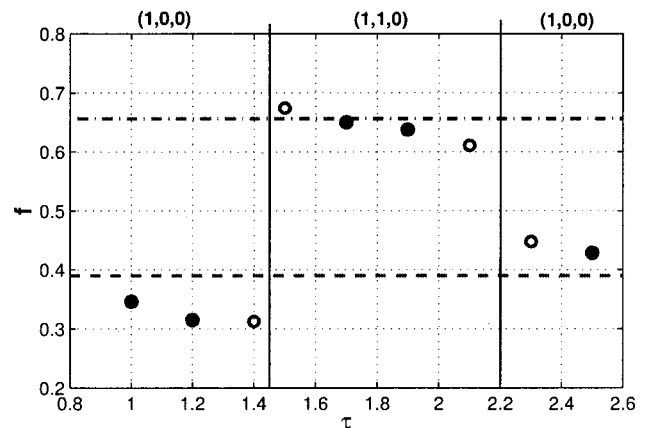


Fig. 13 Frequencies and corresponding mode types for different delay times. (●) Unstable mode, (○) stable mode. Acoustical eigenfrequencies of the (1,0,0)-mode (---) and (1,1,0)-mode (-.-)

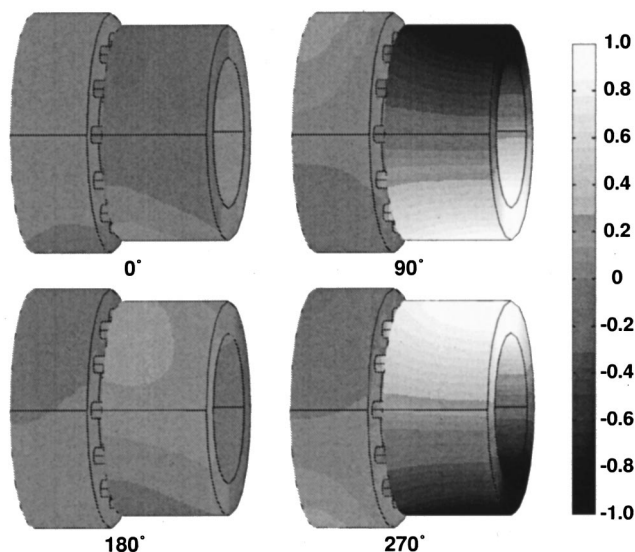


Fig. 14 Normalized pressure distribution at different phase angles for one period of the limit cycle for $\tau=1.9$, $\omega_0=5$, with mean flow

for the two cases which have exhibited the standing mode, the location of the node line has been different. Therefore it is very unlikely that this phenomenon is just the consequence of different initial conditions or other numerical effects.

It makes an important difference if the mode is standing or spinning. This can be seen in Fig. 15 where the amplitudes of the burner exit velocity fluctuations in the limit cycle are given for both cases at each burner. For the spinning mode these amplitudes are nearly the same. The small differences indicate that there also is a small contribution of a standing mode. In the mean flow case, however, the velocity fluctuation amplitudes vary in a wide range from burner to burner. Especially, only few burners exhibit oscillations strong enough to drive the flame into saturation.

In the purely acoustical case each of the two degenerate eigenfunctions is an equally correct solution and any combination of these two solutions could be observed with equal probability. This changes when the coupling to the flames is included as then the distribution of the heat release fluctuation around the circumference is different for different combinations of the two eigensolutions. So it is conceivable that there are factors that support one

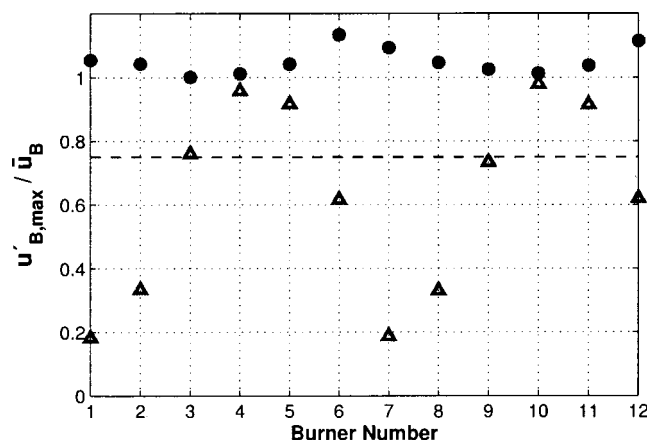


Fig. 15 Amplitudes of the burner exit velocity fluctuations for the cases without (●) and with (△) mean flow ($\tau=1.9$, $\omega_0=5$). Limit for onset of saturation (— — —).

solution more than the other. Here, the mean flow seems to play such an role. This is remarkable as the mean flow Mach numbers appeared to be too low to influence the pure acoustics. So even if we should be careful to generalize this result it must be concluded that *small effects* may decide whether the standing or spinning mode appears. This can be an important issue for active instability control, as adapting the controller to just one of the two modes might easily prove to be not sufficient.

Conclusions and Outlook

We have presented an approach that allows the simulation of combustion instabilities in an annular combustor in the time domain. It is based on the coupling of a nonlinear flame model, describing the heat release rate at each burner, with a linear wave equation for the acoustics, including the effects of a uniform mean flow. The model has proven to be able to predict the growth of an initial perturbation into a finite-amplitude limit cycle oscillation as well as the ultimate decay of any disturbances, depending on the parameters determining the stability of the combustor. To establish a limit cycle, it is necessary to consider both the saturation effects in the flame and the losses appearing in the system.

The results of the simulations of a model annular combustor show that, for the simple heat release model employed here, the delay time plays a major role for the stability of the system and the mode types and frequencies observed. Circumferential modes can occur both in a standing and a spinning form which means a considerable difference regarding the behavior of the individual burners. Here, the mean flow seems to influence which form appears. Otherwise the mean flow does not show any noticeable effect on the stability of the combustor and the frequencies of the oscillations.

The heat release model is a major factor for the results of the calculations. Future attempts to use simulations to predict the behavior of real combustors therefore crucially rely on correct flame models. It is especially important to identify which acoustic quantities at which locations influence the heat release rate fluctuations. The effects of a spatially and temporally varying heat release, caused by a delay time distribution, for example, also must be included. These issues hence should be in the focus of experimental and numerical work dealing with flame response.

In our approach so far we do not consider entropy waves, non-uniform mean flow and loss mechanisms other than sound transmission at the boundaries. It will be part of our future work to extend the technique so that these effects can be included.

We believe that the time domain simulation of combustion instabilities based on our approach can bridge the gap between direct numerical simulations (CFD) which are still computationally out of reach for annular combustors, and low-order models which rely on a correct description of the modal coupling. It can be used as a validation tool for other modeling approaches. The method, moreover, has the potential to be extended to realistic cases and to serve as a design tool in the development of annular combustors.

Acknowledgment

This work has been strongly motivated by a project funded by Siemens Westinghouse and the German Federal Government in the framework of AG Turbo II. We would like to thank Siemens Westinghouse for the valuable discussion within this project.

Nomenclature

- a = velocity of sound
- a = coefficient (FEMLAB)
- c = coefficient (FEMLAB)
- D = mean combustor diameter
- d_a = coefficient (FEMLAB)
- f = frequency
- f = coefficient (FEMLAB)
- g = coefficient (FEMLAB)

h = coefficient (FEMLAB)
 i = imaginary unit
 j = burner number
 K = constant in boundary condition
 L = length
 $\mathbf{n}$ = outward normal vector
 P = thermal power
 p = pressure
 $\dot{q}$ = volumetric heat release rate
 q = coefficient (FEMLAB)
 r = coefficient (FEMLAB)
 T = period
 T = temperature
 t = time
 $\mathbf{u}$ = velocity vector $(u, v, w)^T$
 $\mathbf{u}$ = solution vector (FEMLAB)
 $\mathbf{x}$ = position vector $(x, y, z)^T$
 Y = characteristic impedance

Greek

α = coefficient (FEMLAB)
 β = coefficient (FEMLAB)
 γ = coefficient (FEMLAB)
 η = heat release model parameter
 κ = ratio of specific heats
 λ = Lagrange multiplier (FEMLAB)
 ρ = density
 σ = heat release model parameter
 τ = delay time
 Ω = domain (FEMLAB)
 $\partial\Omega$ = boundary of domain (FEMLAB)
 ω = angular frequency ($2\pi f$)

Superscripts

T = transpose
 $\bar{}$ = mean quantity
 $\prime$ = fluctuating quantity
 $\hat{}$ = complex quantity
 $\sim$ = complex quantity

Subscripts

0 = characteristic quantity
 1 = plenum section
 2 = combustor section
 B = burner section
 in = plenum inlet
 n = outward normal direction
 out = combustor exit
 ∞ = environment

References

- [1] Crocco, L., and Cheng, S. I., 1956, "Theory of Combustion Instability in Liquid Propellant Rocket Motors," AGARD Monograph No. 8.
- [2] Sattelmayer, T., 2003, "Influence of the Combustor Aerodynamics on Combustion Instabilities From Equivalence Ratio Fluctuations," ASME J. Eng. Gas Turbines Power, **125**, pp. 11–19.
- [3] Culick, F. E. C., 1971, "Non-Linear Growth and Limiting Amplitude of Acoustic Oscillations in Combustion Chambers," Combust. Sci. Technol., **3**, pp. 1–16.
- [4] Culick, F. E. C., 1994, "Some Recent Results for Nonlinear Acoustics in Combustion Chambers," AIAA J., **32**(1), pp. 146–169.
- [5] Peracchio, A. A., and Proscia, W. M., 1998, "Nonlinear Heat-Release/Acoustic Model for Thermoacoustic Instability in Lean Premix Combustors," ASME Paper No. 98-GT-269.
- [6] Akamatsu, S., and Dowling, A. P., 2001, "Three Dimensional Thermoacoustic Oscillation in a Premix Combustor," ASME Paper No. 2001-GT-0034.
- [7] Murota, T., and Ohtsuka, M., 1999, "Large-Eddy Simulation of Combustion Oscillation in the Premixed Combustor," ASME Paper No. 99-GT-274.
- [8] Brookes, S. J., Cant, R. S., and Dowling, A. P., 1999, "Modelling Combustion Instabilities Using Computational Fluid Dynamics," ASME Paper No. 99-GT-112.
- [9] Hantschk, C.-C., and Vortmeyer, D., 2002, "Numerical Simulation of Self-Excited Combustion Oscillations in a Non-Premixed Burner," Combust. Sci. Technol., **174**, pp. 189–204.
- [10] Wake, B. E., Choi, D., and Hendricks, G. J., 1996, "Numerical Investigation of Pre-Mixed Step-Combustor Instabilities," AIAA Paper No. AIAA 96-0816.
- [11] Dowling, A. P., 1997, "Nonlinear Self-Excited Oscillations of a Ducted Flame," J. Fluid Mech., **346**, pp. 271–290.
- [12] Walz, G., Krebs, W., Hoffmann, S., and Judith, H., 1999, "Detailed Analysis of the Acoustic Mode Shapes of an Annular Combustion Chamber," ASME Paper No. 99-GT-113.
- [13] Walz, G., Krebs, W., Flohr, P., and Hoffmann, S., 2001, "Modal Analysis of Annular Combustors: Effect of Burner Impedance," ASME Paper No. 2001-GT-0042.
- [14] Krüger, U., Hüren, J., Hoffmann, S., Krebs, W., and Bohn, D., 1999, "Prediction of Thermoacoustic Instabilities With Focus on the Dynamic Flame Behavior of the 3A-Series Gas Turbine of Siemens KWU," ASME Paper No. 99-GT-111.
- [15] Krüger, U., Hüren, J., Hoffmann, S., Krebs, W., Flohr, P., and Bohn, D., 2001, "Prediction and Measurement of Thermoacoustic Improvements in Gas Turbines With Annular Combustion Systems," ASME J. Eng. Gas Turbines Power, **123**, pp. 557–566.
- [16] Krebs, W., Walz, G., and Hoffmann, S., 1999, "Thermoacoustic Analysis of Annular Combustor," AIAA Paper No. AIAA 99-1971.
- [17] Stow, S. R., and Dowling, A. P., 2001, "Thermoacoustic Oscillations in an Annular Combustor," ASME Paper No. 2001-GT-0037.
- [18] Evesque, S., and Polifke, W., 2002, "Validation of Low-Order Acoustic Modelling for Annular Combustors," ASME Paper No. GT-2002-30064.
- [19] Ffowcs Williams, J. E., 1982, "Sound Sources in Aerodynamics—Fact and Fiction," AIAA J., **20**(3), pp. 307–315.
- [20] Polifke, W., Paschereit, C. O., and Döbeling, K., 2001, "Constructive and Destructive Interference of Acoustic and Entropy Waves in a Premixed Combustor With a Choked Exit," Int. J. Acoust. Vib., **6**(3), pp. 135–146.
- [21] Flohr, P., Paschereit, O. C., van Roon, B., and Schuermans, B., 2001, "Using CFD for Time-Delay Modeling of Premix Flames," ASME Paper No. 2001-GT-0376.
- [22] COMSOL AB, "FEMLAB Reference Manual," 2000.
- [23] Harper, J., Johnson, C., Neumeier, Y., Lieuwen, T. C., and Zinn, B. T., 2001, "Experimental Investigation of the Nonlinear Flame Response to Flow Disturbances in a Gas Turbine Combustor," AIAA Paper No. AIAA-01-0486.
- [24] Chu, B.-T., 1956, "Stability of Systems Containing a Heat Source—The Rayleigh Criterion," NACA Research Memorandum NACA RM 56D27.

On-Line Fuel Deoxygenation for Coke Suppression

L. J. Spadaccini

H. Huang

United Technologies Research Center,
411 Silver Lane,
East Hartford, CT 06108

Fuel deoxygenation is being developed as a means for suppressing autoxidative coke formation in aircraft fuel systems, thereby increasing the exploitable cooling capacity of the fuel, enabling major increases in engine operating temperature and cycle efficiency. Reduced maintenance is an added benefit. A prototype membrane filter module for on-line removal of dissolved oxygen, which would otherwise react to form coke precursors, was constructed and successfully demonstrated. The fuel flows over the membrane, while oxygen diffuses through it at a rate that is proportional to the difference in oxygen partial pressures across the surface. Tests were conducted over a range of fuel flow rates (residence times) and temperatures. The filter was operated with air-saturated jet fuel for several hours at a steady-state condition, verifying the capability to remove essentially all of the dissolved oxygen (to <1 ppm) and proving the viability of the concept. A convincing demonstration of coke suppression was performed when air-saturated (normal) and deoxygenated jet fuels were tested in a standard ASTM heated tube apparatus at wall temperatures as high as 850°F. With deoxygenated fuel, there was a dramatic reduction (more than an order of magnitude) in coke deposition relative to air-saturated Jet A, which will allow the maximum fuel temperature to be increased by more than 200°F, doubling the available heat sink. Moreover, deoxygenated Jet A was shown to perform as well as JP-7, the Air Force's highest thermal stability fuel. An analytical model for oxygen permeation through the membrane was formulated, and used in conjunction with the test data to estimate the filter size required for a practical (i.e., low-volume/high flow rate) deoxygenator. [DOI: 10.1115/1.1582497]

Introduction

Thermal management is a design driver for advanced gas turbine engines. Increasing cycle temperature increases engine performance and cycle efficiency. However, improved materials for enabling high-temperature operation are not available. Fuel is currently used as a heat sink for the engine oil system, but its cooling capacity is limited by a temperature constraint necessary to preclude the formation of coke deposits. Therefore, it is essential to develop a method to suppress coke formation and, thereby, significantly increase the available heat sink and permit extended utilization of the fuel for component and system cooling (e.g., turbine, compressor, and avionics) in advanced engines and sustained flight at high Mach number. Over the past 40 years there have been many attempts by researchers to suppress coking, [1,2], but, with the exception of JP-8+100 additives, [3], that allow operation up to 425°F, they have proven to be unsuccessful or impractical.

The major factor contributing to coke deposition at temperatures up to approximately 700°F is oxygen that dissolves into the fuel when it comes in contact with air, [4]. When air-saturated fuel is heated above approximately 300°F, the dissolved oxygen reacts to form free-radical species (coke precursors) that initiate and propagate other autoxidation reactions, leading to deposit formation. These reaction paths become insignificant when the concentration of the dissolved oxygen is reduced from its ambient saturation level of 70 ppm to approximately 1 ppm (i.e., deoxygenated), [1]. At temperatures above approximately 900°F, the deposition mechanism is characterized by pyrolysis, wherein chemical bonds are broken and large alkanes are converted into smaller alkanes, alkenes, and some hydrogen, [5]. Since pyrolysis requires much higher temperatures than autoxidation, reducing the

oxygen content of fuel will allow it to be heated significantly before thermal decomposition begins, increasing the cooling capacity. Therefore, fuel deoxygenation should dramatically reduce autoxidative coke deposition, and make it possible to realize the thermal stability goals for JP-8+225/JP-900, [6].

This study deals with the development of a small, practical membrane-based fuel deaerator for on-line removal of dissolved oxygen that would otherwise react and form the precursors for autoxidative coke. A permeable membrane is a selective barrier that permits the separation of certain species in a fluid by diffusion or sorption-diffusion mechanisms. In general, the membrane structure consists of an ultra-thin coating that has the requisite separation properties, and a micro-porous polymer support that provides strength. In the present representation (see Fig. 1), the oxygen molecules in jet fuel dissolve into the membrane and then diffuse across it, driven by the difference of oxygen partial pressure (chemical potential or driving force), while the hydrocarbon molecules are unaffected and pass over it. Previous attempts at deoxygenating fuel have included sparging with nitrogen, and the use of molecular-sieve adsorbents and chemical reducing agents. These approaches have proven impractical for aircraft applications because they are costly, heavy and bulky, or even dangerous, [7]. Some aircraft use an on-board inert gas generator system (OBIGGS) to reduce the oxygen concentration in the fuel tank below the flammability limit (~9 vol. %). However, it is unlikely that a similar system could be used to lower the oxygen concentration dissolved in the fuel to approximately 1 ppm.

Experimental Approach

An experimental apparatus for measuring the oxygen removal rate from jet fuel by membrane permeation has been developed and is illustrated schematically in Fig. 2. The rig consists of a fuel aeration system, a fuel pump, a preheater, a membrane module, and an on-line oxygen sensor. Although the membrane is envisioned as being very thin to facilitate gas diffusion (e.g., micron

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Amsterdam, The Netherlands, June 3–6, 2002; Paper No. 2002-GT-30071. Manuscript received by IGTI, December 2001, final revision, March 2002. Associate Editor: E. Benvenuti.

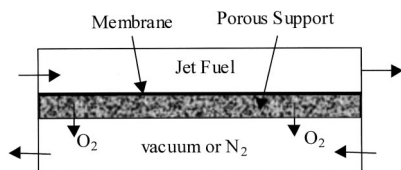


Fig. 1 Supported oxygen-permeable polymer membrane

thickness), for convenience, concept feasibility experiments were conducted using several closely sized commercially available thin-wall tubes made of the membrane material.

In a fuel deoxygenation test, air-saturated jet fuel was metered through the membrane module at a predetermined and controllable pressure, temperature, and flow rate. After leaving the module, the fuel was cooled to ambient temperature (to eliminate measurement uncertainties) and the oxygen concentration determined on-line using a polarographic-type oxygen sensor. The membrane module consisted of a polymeric membrane tube (through which the fuel flowed) installed in a cylindrical shell that was either evacuated or purged with nitrogen, to provide the driving force (oxygen partial pressure difference) for oxygen removal. Tests were conducted in a single-pass flow arrangement over a range of fuel flow rates (residence times) and temperatures. The maximum fuel temperature was limited to 240°F to preclude oxygen depletion due to thermal reactions. This behavior was verified by substituting a stainless-steel tube for the membrane tube and demonstrating that at 240°F the oxygen concentration at the outlet was unchanged from the inlet (i.e., 100% saturated). To verify the

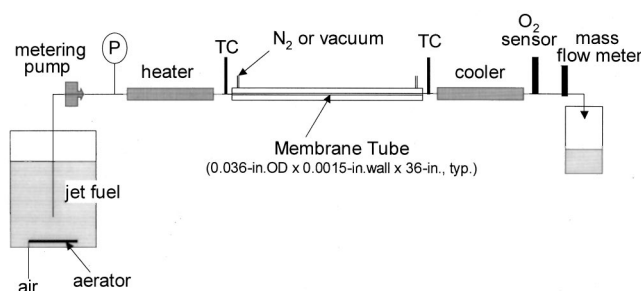


Fig. 2 Membrane test apparatus

oxygen sensor accuracy/repeatability after each test, the fuel was re-aerated to demonstrate measurement of 100% saturation.

An initial series of tests was conducted to measure the permeability of a 0.036-in. OD×0.0015-in. wall membrane tube to gaseous oxygen. The membrane is essentially impermeable to liquids and organic vapors, and the gas permeability increases as the size of the molecule decreases. The oxygen permeability of the membrane was determined by pressurizing the tube with oxygen, and measuring the amount of gas that diffused through the tube wall into the outer shell. Oxygen permeation rates were determined using bubble-type flowmeters for measuring very low flow rates. The pressure and temperature of the oxygen feed were varied, and the outer shell was maintained at atmospheric pressure. The feed and permeate sides of the module were purged with oxygen prior to each permeation test. The results are presented in Fig. 3, and show that the oxygen permeability of the polymer membrane tube increases linearly with the oxygen pressure difference across it, and also increases significantly with increasing temperature.

Another important factor that will influence the oxygen removal rate is the membrane thickness. Additional oxygen-gas permeability tests were conducted to evaluate the effect of membrane thickness using membrane tubing with wall thickness of 0.001 in. and 0.003 in. The results, shown in Fig. 4, for ambient temperature (70°F) indicate that the oxygen gas diffusion through the membrane increases rapidly (nearly exponentially) with decreasing thickness, and emphasizes the desirability of the membrane being relatively thin.

The influence of fuel temperature on the rate of oxygen removal is shown in Fig. 5, for a 0.003-in.-thick membrane tube. At ambient temperature, the oxygen concentration in air-saturated JP-8 fuel was reduced by 70% (to approximately 20 ppm) in a single pass through the 3-ft long, 0.040-in. OD polymeric membrane tube. At a fuel temperature of 240°F (a typical aircraft main fuel pump outlet temperature), more than 90% of the dissolved oxygen was removed for the same residence time (volume flow rate). That is, the oxygen concentration was reduced from approximately 70 ppm to a level below 5 ppm in a residence time of 25 seconds. This rate of deoxygenation corresponds to a space velocity of 145/hr, where

$$SV = \frac{\text{volume flow rate}}{\text{reactor volume}} = \frac{1}{\text{residence time}} \quad (1)$$

In an aircraft fuel system, higher space velocities will be required to minimize the filter volume. Reducing the membrane thickness and increasing the surface area will enhance oxygen

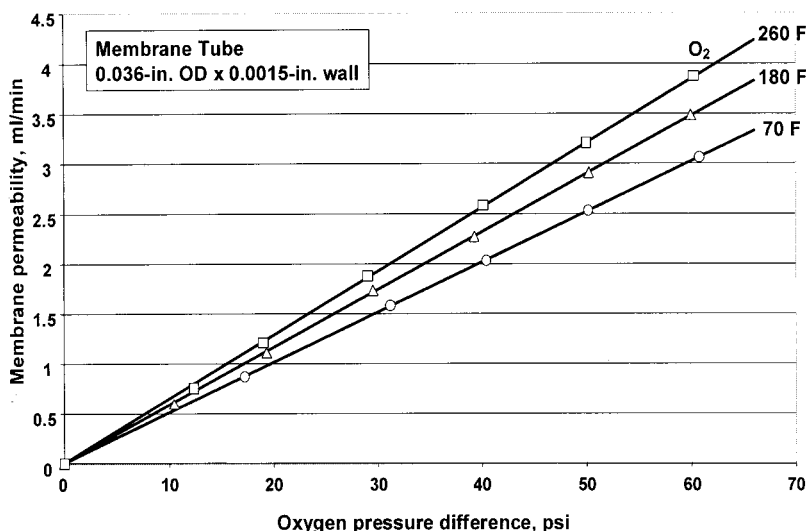


Fig. 3 Influence of oxygen temperature on membrane permeability

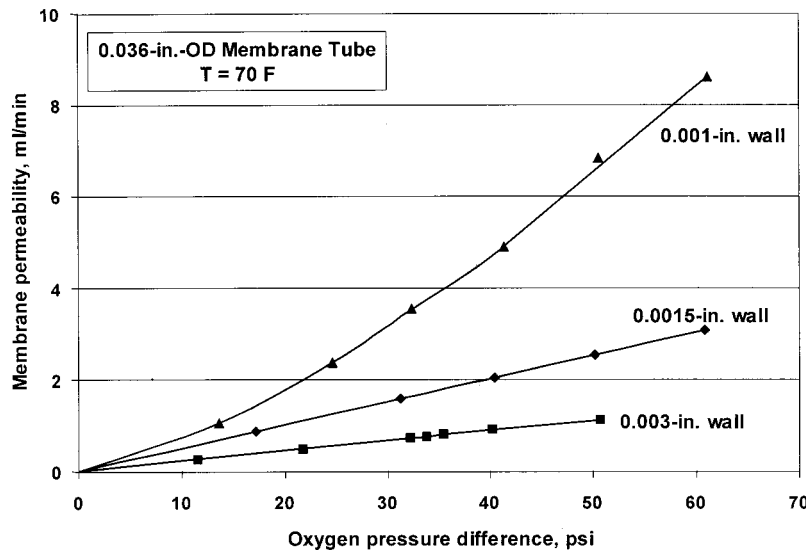


Fig. 4 Influence of membrane thickness on oxygen permeability

diffusion and facilitate operation at much higher space velocity. Fuel heating enhances performance by increasing the permeability of the membrane, and by decreasing the solubility of oxygen in fuel, thereby increasing the driving force across the membrane. Increasing the pressure of the air-saturated fuel will not produce a significant benefit because it will not change the oxygen partial pressure (which is established in the tank) or significantly increase the permeability of the membrane.

Diffusion of the oxygen molecules dissolved in liquid fuel to the membrane surface can be the rate-determining or rate-limiting step in the fuel deoxygenation process, depending on the oxygen diffusion rate through the membrane. When an effective and thin membrane is used, the overall deoxygenation rate may be controlled by the bulk diffusion of the oxygen molecules in fuel. This effect can be quantitatively described by the oxygen mass transfer coefficient,

$$f_{MT} = \frac{\text{overall oxygen removal rate}}{\text{oxygen diffusion rate through membrane}}, \quad (2)$$

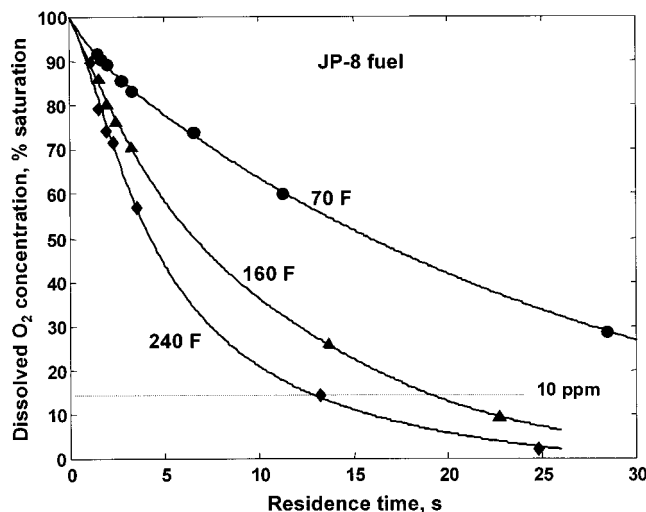


Fig. 5 Influence of fuel temperature on deoxygenation

where the oxygen diffusion through the membrane is determined from the gas permeability tests discussed above. Mass transfer coefficients for flow in a 0.034-in. ID $\times$ 0.003-in. wall membrane tube are shown in Fig. 6, as a function of the fuel flow rate. The results indicate that the mass transfer coefficient increases as the fuel flow rate increases from 50 g/hr ($Re=16$) to about 800 g/hr ($Re=260$). Although the flow is still laminar at 800 g/hr, the bulk diffusion rate of the oxygen molecules in fuel is much faster than the rate through the membrane and, therefore, the mass transfer coefficient approaches 1.0, indicating that there was no limitation caused by mass transfer. Mass transfer in the fuel did limit the overall deoxygenation rate when the fuel flow rate was less than 800 g/hr (i.e., $Re < 260$).

Concept Demonstration

A prototype fuel deoxygenator module was constructed for measuring the oxygen removal rate from jet fuel, and tests were conducted over a range of fuel flow rates (residence times) and temperatures. The module consists of a spool of 0.040-in. ID $\times$ 0.003-in. wall oxygen-permeable membrane tubing installed in a cylindrical container. The container is purged with a low flow rate of nitrogen to create an oxygen-free atmosphere around the tube. Fuel flows inside the tubing, while the dissolved oxygen diffuses through the wall at a rate that is proportional to the difference in the oxygen partial pressures (driving force) across the surface.

Several tests were performed to demonstrate the capability to deoxygenate air-saturated JP-8 fuel to very low concentration levels (< 1 ppm) in a single-pass on-line flow arrangement. The tests were conducted with aerated fuel at ambient temperature, a much more difficult condition than occurs in an aircraft fuel system. (In Fig. 5, it was shown that the deoxygenation rate increases with increasing temperature.) Typical results, shown in Fig. 7, verify the rapid removal of essentially all of the dissolved oxygen from the fuel, beginning shortly after initiating the nitrogen purge gas flow. The fuel deoxygenator module was operated at a steady-state condition for more than one hour before terminating the nitrogen bleed and initiating an air purge, thereby ending the deoxygenation process (as indicated by the return of the oxygen sensor to the air-saturated-fuel starting condition). This result clearly demonstrates the feasibility of the on-line deoxygenation concept. The fuel pressure loss through the membrane filter can be minimized

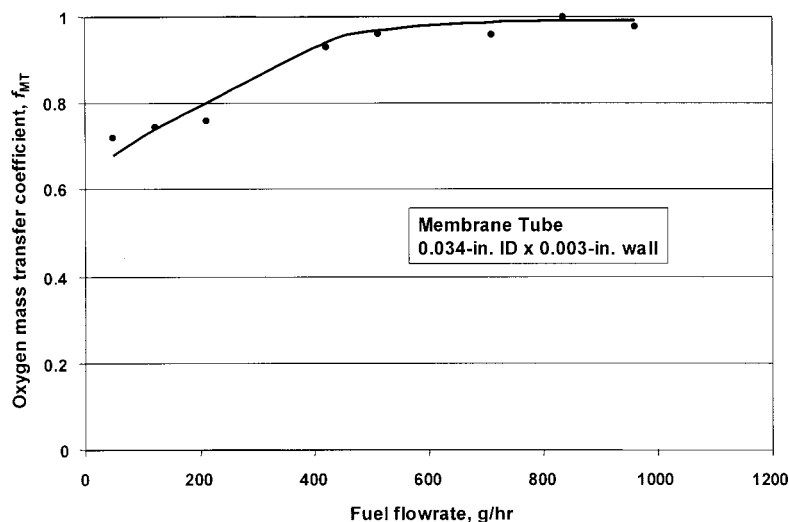


Fig. 6 Influence of mass diffusion in fuel on deoxygenation

in a practical design by interchanging the fuel and nitrogen flows, with the fuel being directed over and around a bundle of membrane tubes (enhancing mass transport) and oxygen permeating into the tubes. A candidate design is described below.

To provide a simple demonstration of coke suppression by on-line fuel deoxygenation, a small-scale test apparatus was constructed and tests were performed with typical jet fuels. The fuel deoxygenator module was used in conjunction with an ALCOR hot liquid process simulator (HLPS) to carry out a series of standard ASTM jet fuel thermal oxidation stability tests (JFTOT procedure, ASTM Standard Method D 3241). The JFTOT is a pass-fail tester for specification purposes, and not a precision instrument for rating small differences in carbon deposition among fuels. It is used here to reveal potentially large differences in coke formation resulting from fuel deoxygenation.

Heated-tube JFTOT procedure tests were conducted with air-saturated fuel (baseline) and fuels deoxygenated on-line using the membrane module. The dissolved oxygen concentration was monitored continuously and observed to remain constant at either the saturation or the near zero level (shown in Fig. 7). Because the tube is heated resistively (i.e., by an electrical current), the axial temperature profile is very steep. Type 316 stainless steel tubes

were substituted for conventional aluminum tubes to permit operation at high temperatures and quantification of the surface deposit by carbon burn off using a LECO RC-412 carbon determinator (precision of $\pm 2\%$). Evaluations were performed with Jet A, JP-8, and JP-7 in five-hour tests at maximum tube temperatures of 635–860°F and fuel flow rate of 3 ml/min, conditions that are much more severe than the normal JFTOT procedure. These fuels and run conditions were used previously in the development program for JP-8+100, and an extensive database is available for comparison of results, [8].

The test conditions and results are summarized in Table 1. They were performed in the order shown (including carbon burn off), i.e., alternating between air-saturated and deoxygenated fuel. With deoxygenated fuel, there was a dramatic reduction (more than an order of magnitude) in coke deposition on the surface and in particulate suspended in the bulk flow. For the deoxygenated fuel tests, no increase in pressure loss was measured across an in-line test filter located at the exit of the heater tube. Surface deposition results presented in Fig. 8 demonstrate that deoxygenation suppresses coke formation, independent of fuel type (cf., Jet A, JP-8, and JP-8+100) or batch (cf., 91-POSF-2827 and 96-POSF-3219). The data in Fig. 9 indicate that the deoxygenator is effective in suppressing autothermal reactions up to 860°F peak wall temperatures, confirming the initial hypothesis. Furthermore, deoxygenated Jet A performed just as well as JP-7, a special high thermal stability fuel that is highly processed and much more costly (see Fig. 10). Therefore, it can be inferred that deoxygenation could

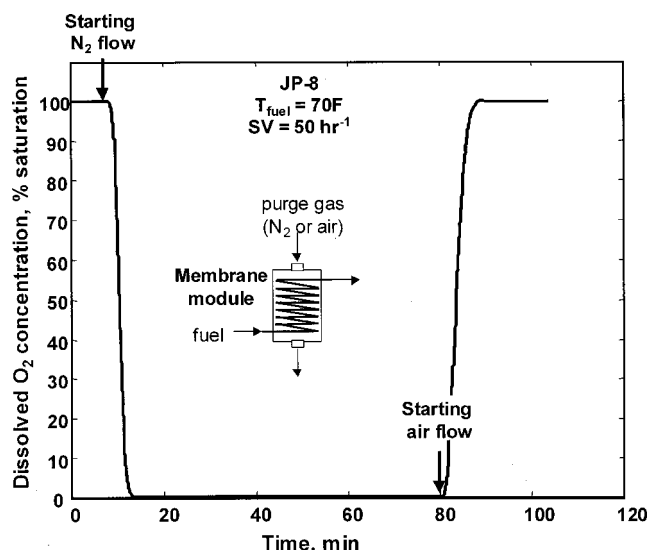


Fig. 7 Demonstration of single-pass fuel deoxygenation

Table 1 Jet fuel thermal oxidation tests

Fuel	Running Mode	Peak Wall Temperature °F	Surface Carbon $\mu\text{g}/\text{cm}^2$	Pressure Drop Across Filter mm-Hg/min.
Jet-A*	air-saturated	635	103	240/105
Jet-A*	deoxygenated	635	4	0/300
Jet-A*	deoxygenated	700	3	1/300
Jet-A*	deoxygenated	860	6	12/300
Jet-A**	air-saturated	635	61	240/160
Jet-A**	deoxygenated	635	4	0/300
JP-8	air-saturated	635	92	1/300
JP-8	deoxygenated	700	4	0/300
JP-7	air-saturated	700	5	0/300
JP-7	air-saturated	840	4	1/300
JP-8+100	air-saturated	635	25	0/300
JP-8+100	deoxygenated	635	1	0/300

*91-POSF-2827

**96-POSF-3219

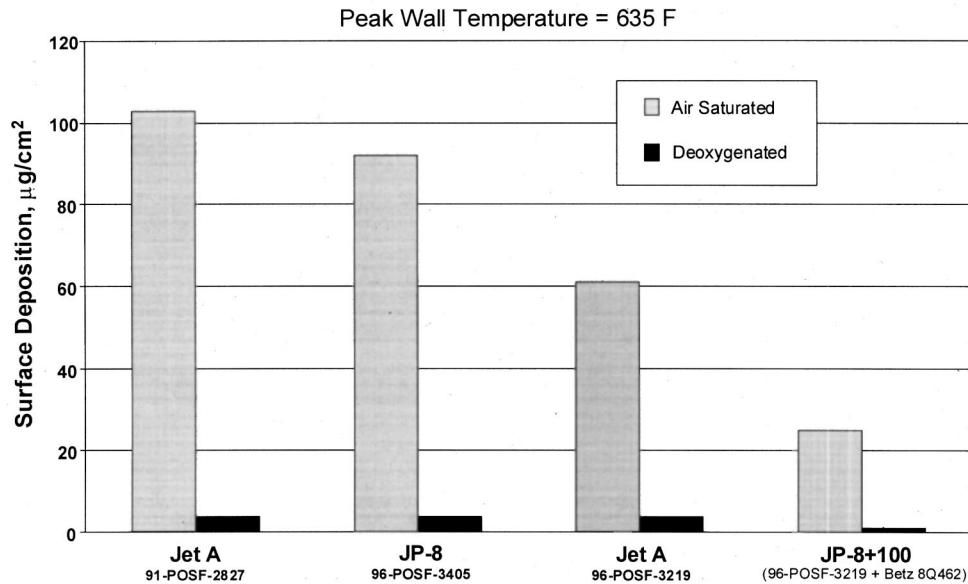


Fig. 8 Deoxygenation suppresses coke formation, independent of fuel

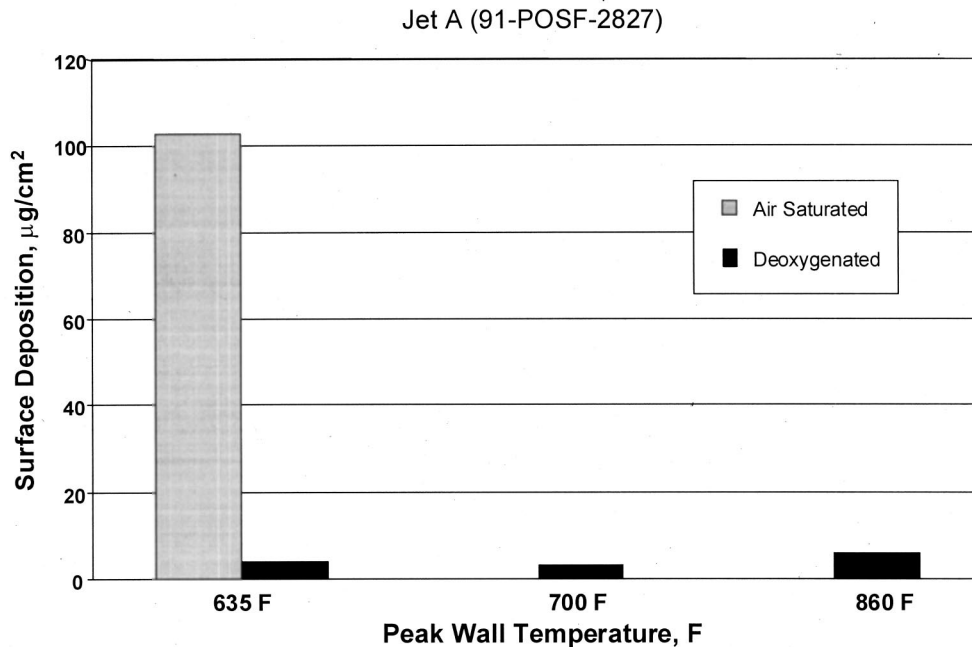


Fig. 9 Deoxygenation is effective up to 860°F peak wall temperature

make possible increasing the maximum allowable temperature of Jet A to that of JP-7, or from 325°F to 550°F, more than doubling the available heat sink. The membrane module was operated for approximately 40 hours with no change in performance. Clearly, additional testing in simulator rigs and engines is required to validate the concept and establish the level of deoxygenation necessary for coke mitigation.

Modeling

Oxygen permeation rate through a membrane is governed by the fundamental rate equation, [9]:

$$\frac{dn_{O_2}(\text{out})}{dt} = \frac{k(\Delta P_{O_2})A}{\delta} \quad (3)$$

where k is the oxygen permeability of the membrane, ΔP_{O_2} the oxygen partial pressure difference (driving force) across the membrane, A the membrane surface area, δ the membrane thickness, and $dn_{O_2}(\text{out})/dt$ the oxygen permeation rate through the membrane. A mechanism-based kinetic model (function of time, temperature, pressure, and concentration) for oxygen permeation through the membrane was derived from Eq. (3) and used in conjunction with the test data to estimate the filter size required for a practical (i.e., low volume/high flow rate) deoxygenator.

Oxygen removal efficiency (η_{OR}) is defined as

$$\eta_{OR} = 1 - \frac{C_{O_2}}{C_{O_2(0)}} \quad (4)$$

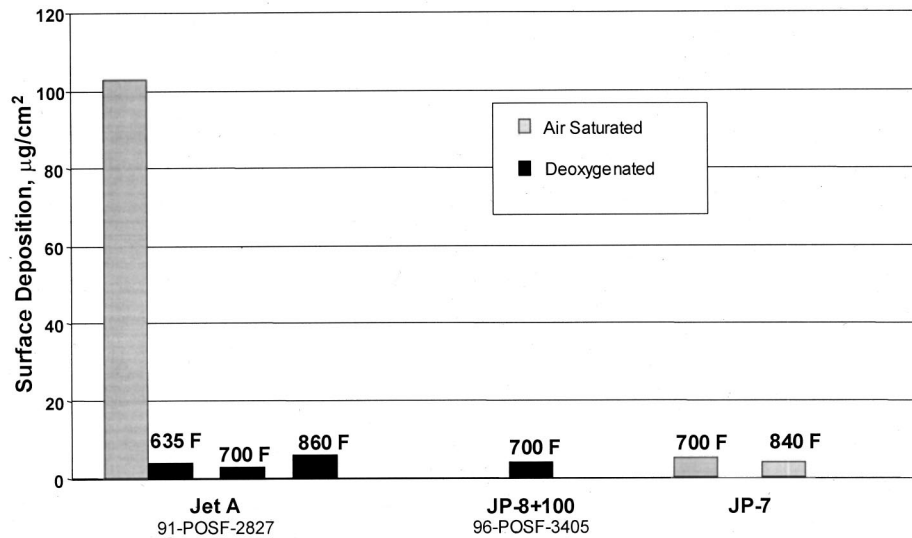


Fig. 10 Deoxygenated fuel performs as well as JP-7

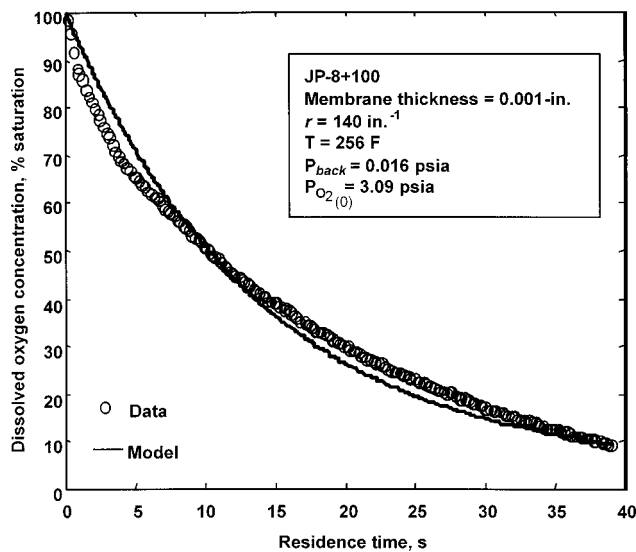


Fig. 11 Kinetic model validation

where $C_{O_{2(0)}}$ is the initial concentration of oxygen dissolved in jet fuel at ambient conditions, and C_{O_2} is the final concentration of dissolved oxygen. The oxygen removal efficiency is primarily determined by the partial pressure of oxygen on the backside of the membrane (P_{back}), the membrane permeate diffusion coefficient (λ), the membrane surface-area/volume ratio (r), the oxygen mass transfer coefficient (f_{MT}), and the residence time (t). The performance equation developed from Eqs. (3) and (4) is given by

$$\eta_{OR} = 1 - \left[\left(1 - \frac{P_{back}}{P_{O_{2(0)}}} \right) \cdot \exp \left(- \frac{f_{MT} r \lambda}{\gamma} t \right) + \frac{P_{back}}{P_{O_{2(0)}}} \right] \quad (5)$$

In this equation, $P_{back} = 0.21 \cdot P_{vacuum}$, or 0 for N_2 purge, $\lambda =$ diffusion coefficient (i.e., permeability/thickness), and $\gamma = P_{O_2}/C_{O_{2(0)}}$ (for Jet A, JP-8, and JP-8+100 $\gamma = 0.0127$). The key kinetic parameter, the oxygen diffusion coefficient (λ) through the membrane, is determined from the membrane permeability test data shown in Fig. 3. The mass transfer coefficient of oxygen (f_{MT}) in liquid fuel is a constant for a particular operating condition and fuel flow configuration. As a demonstration of model validation, a comparative plot of test data against model predictions is illustrated in Fig. 11, and shows very good agreement. Under these conditions, the fuel flow was highly turbulent.

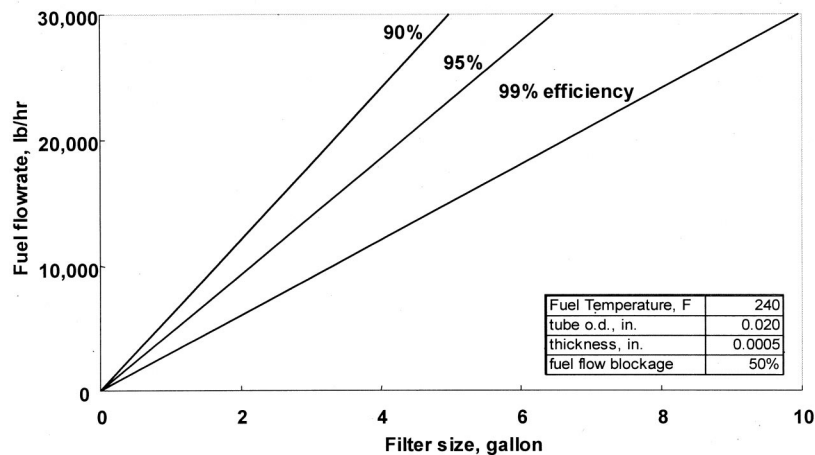


Fig. 12 Fuel deoxygenation filter sizing

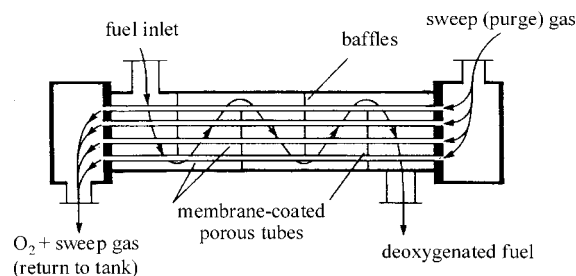


Fig. 13 Fuel deoxygenator design concept

Therefore, the mass transfer coefficient was assumed to be 1. This assumption was validated experimentally. To develop a more comprehensive design tool, fundamental parameters for estimating the mass transfer coefficient of oxygen in fuel are required.

The membrane permeation model described by Eq. (5) was used as a conceptual design tool. Simulations were carried out to size an on-line fuel deoxygenator for the maximum ($\sim 30,000$ lb/hr at takeoff) flow rate requirements of a typical commercial aircraft engine, and for different levels of fuel deoxygenation (i.e., 90 to 99%, or 7 to 0.7 ppm). A 50% fuel flow (open) area is assumed for the optimum combination of residence time and membrane surface area, corresponding to the minimum component volume. The results, presented in Fig. 12, illustrate that a deoxygenator of 10-gallon volume would be adequate, and a smaller volume sufficient if the percentage oxygen removal requirement (efficiency) could be relaxed. Note that the filter size is very dependent on the level to which the fuel must be deoxygenated, since the driving force decreases exponentially as oxygen is removed from the fuel. A determination of the optimum level of fuel deoxygenation was beyond the scope of this study. Alternatively, the deoxygenator might more appropriately be sized for the cruise condition, where engine temperatures are high, fuel flow rate is lower than during acceleration, and flight time is longest.

System Design/Development

The membrane deoxygenator can be configured like a shell-and-tube heat exchanger to maximize the surface area of the membrane contacted by the fuel and minimize the volume of the component. As illustrated in Fig. 13, the fuel will flow through the shell, over and around many micro-porous tubes that have been coated on the outside with an ultra-thin membrane layer. The inclusion of flow baffles in the shell enhances the diffusion of oxygen to the surface by forcing the fuel to follow a lengthy, tortuous path over the membrane. Dissolved oxygen will permeate into the tubes from the shell, driven by the difference in chemical potential across the membrane.

Control of the oxygen partial pressure on the backside of the membrane can be accomplished by the use of a vacuum pump or by purging with a low flow rate of a gas that does not contain oxygen (a sweep gas). Because no more than 70-ppm oxygen is being removed from the fuel, only a small vacuum pump would be required (less than 1/4 HP for a typical commercial aircraft). For operating safety, any fuel (liquid or vapor) that may leak through the membrane will be returned to the fuel storage tank. Alternatively, the sweep gas can be precharged and recirculated in a closed loop, and oxygen transferred from the fuel can be removed by an adsorbent that would be replaced periodically. Also, the pressure of the sweep gas could be controlled to minimize the stress on the membrane.

The next step in the development process is to construct and demonstrate a reduced-scale prototype, and to validate the design and fabrication methodology through extended duration testing in fuel system simulators. A detailed system design/integration and the construction of a full-size component would follow, leading to a final phase to fabricate a full-scale component and demonstrate performance in an engine test.

Conclusion

The results demonstrate the feasibility of a novel on-line fuel deoxygenation concept, and the potential for improving the fuel thermal-oxidative stability and increasing the cooling capacity. The principal conclusions are:

- on-line removal of dissolved oxygen from jet fuel using a membrane filter is a feasible method for significantly increasing the usable heat sink of the fuel. The results suggest that practical size deoxygenators may be designed for use in aircraft systems.
- fuel deoxygenation is very effective in suppressing autooxidative coke formation, making it possible to increase the maximum allowable temperature and more than double the available heat sink. The membrane filter is capable of deoxygenating the fuel to a level below that at which significant coking occurs, and should reduce maintenance in aircraft fuel systems.
- the key variables controlling fuel deoxygenation are the difference in oxygen partial pressure across the membrane, the membrane thickness, and oxygen diffusion in the fuel. Thin membranes, operation at elevated temperature, and turbulent flows enhance performance.
- fuel heating increases the permeability of the membrane and the chemical potential across it.
- the mechanism-based kinetic model for oxygen permeation through a membrane is useful for designing a practical deoxygenator component.

Acknowledgments

This paper is based on research supported by the United Technologies Corporation, and work performed for the Air Force Research Laboratory, Propulsion Directorate under Contract F33615-97-D-2784, administered by Dr. Tim Edwards of the Fuels Branch, Turbine Engine Division. The authors gratefully acknowledge the test support provided by Mr. David McHugh.

References

- [1] Hazlett, R. N., 1991, *Thermal Oxidation Stability of Aviation Turbine Fuels*, ASTM Monograph 1, American Society for Testing and Materials, Philadelphia, PA.
- [2] Spadaccini, L. J., Sobel, D. R., and Huang, H., 2001, "Deposit Formation and Mitigation in Aircraft Fuels," *ASME J. Eng. Gas Turbines Power*, **123**, pp. 741–746.
- [3] Heneghan, S. P., Zabarnick, S., Ballal, D. R., and Harrison, W. E., 1996, "JP-8+100: Development of a Thermally Stable Jet Fuel," *ASME J. Energy Resour. Technol.*, **118**, pp. 170–179.
- [4] Taylor, W. F., 1974, "Deposit Formation From Deoxygenated Hydrocarbons. General Features," *Ind. Eng. Chem., Prod. Res. Develop.*, **13**, pp. 133–138.
- [5] Sobel, D. R., and Spadaccini, L. J., 1997, "Hydrocarbon Fuel Cooling Technologies for Advanced Propulsion," *ASME J. Eng. Gas Turbines Power*, **119**, pp. 344–351.
- [6] Edwards, T., 1998, "Prospects for JP-8+225, a Stepping Stone to JP-900," AIAA Paper 98-3532.
- [7] Darrah, S., 1988, "Jet Fuel Deoxygenation," Report AFWAL-TR-88-2081, Air Force Aero Propulsion Laboratory, WPAFB, OH.
- [8] Biddle, T., 1999, "JP-8+100: An Improved Thermally Stable JP-8 Fuel for Current and Future Aircraft," Report AFRL-PR-WP-TR-2000-2019, Air Force Aero Propulsion Laboratory, WPAFB, OH.
- [9] Baker, R., 2000, *Membrane Technology and Applications*, McGraw-Hill, New York.

Flashback Limits for Combustion Induced Vortex Breakdown in a Swirl Burner

M. Kröner

J. Fritz

T. Sattelmayer

Lehrstuhl für Thermodynamik,
Technische Universität München,
85747 Garching, Germany

Flame flashback from the combustion chamber into the mixing zone limits the reliability of swirl stabilized lean premixed combustion in gas turbines. In a former study, the combustion induced vortex breakdown (CIVB) has been identified as a prevailing flashback mechanism of swirl burners. The present study has been performed to determine the flashback limits of a swirl burner with cylindrical premixing tube without centerbody at atmospheric conditions. The flashback limits, herein defined as the upstream flame propagation through the entire mixing tube, have been detected by a special optical flame sensor with a high temporal resolution. In order to study the effect of the relevant parameters on the flashback limits, the burning velocity of the fuel has been varied using four different natural gas-hydrogen-mixtures with a volume fraction of up to 60% hydrogen. A simple approach for the calculation of the laminar flame speeds of these mixtures is proposed which is used in the next step to correlate the experimental results. In the study, the preheat temperature of the fuel mixture was varied from 100°C to 450°C in order to investigate influence of the burning velocity as well as the density ratio over the flame front. Moreover, the mass flow rate has been modified in a wide range as an additional parameter of technical importance. It was found that the quenching of the chemical reaction is the governing factor for the flashback limit. A Peclet number model was successfully applied to correlate the flashback limits as a function of the mixing tube diameter, the flow rate and the laminar burning velocity. Using this model, a quench factor can be determined for the burner, which is a criterion for the flashback resistance of the swirler and which allows to calculate the flashback limit for all operating conditions on the basis of a limited number of flashback tests. [DOI: 10.1115/1.1582498]

Introduction

Flame flashback from the combustion chamber into the mixing zone is one of the inherent reliability problems of lean premixed combustion. In order to achieve a sufficient margin against flashback under all operating conditions, the prediction of the expected flashback limits is a crucial burner design task. The motivation of the presented study was that reliable predictive methods are not yet available. For this reason, a method is developed which allows the prediction of the flashback limits as a function of the operating conditions on the basis of a characteristic quench factor, which can be determined with a limited number of flashback tests.

In swirl burners flashback can be initiated by the four following causes, [1]:

- flame propagation in the boundary layer,
- turbulent flame propagation in the core flow,
- violent combustion instabilities, and
- combustion induced vortex breakdown CIVB.

The combustion induced vortex breakdown has been identified as a prevailing flashback mechanism in swirl burners without centerbody, [1]. It has been shown that although the vortex breakdown in the isothermal flow is located at the combustion chamber inlet and the axial velocities in the vortex core are large (Fig. 1(a)), the swirling flow can nevertheless break down in the mixing tube due to the interaction with the chemical reaction. The induced recirculation region propagates upstream and the flame tip is attached to the breakdown point during this process (Fig. 1(b)).

This phenomenon discovered in [1] is the reason for the observed upstream flame propagation against a strong main flow in the downstream direction.

Flame vortex interaction as the driving force for rapid flame propagation in swirling flows is well known from studies of flames in small vortex tubes or vortex rings (McCormack et al. [2]). Different physical models have been derived taking the pressure forces and axial momentum flux of the burned gas (Asato et al. [3] and Chomiak [4]¹) or the conservation equations over the flame front and Bernoulli's theorem into consideration (Umemura et al. [5] and Ishizuka et al. [6]²).

In these generic studies, the density ratio over the flame front $\sigma = \rho_u / \rho_b$ and the maximum circumferential velocity $w_{\max}$ of the assumed rankine vortex have been found as the common influencing parameters on the flame speed S_f :

$$S_f = f(\bar{w}_{\max}, \sigma). \quad (1)$$

With increasing $\bar{w}_{\max}$ the flame velocity rises considerably. Whether the flame speed is an increasing or decreasing function of the density ratio σ depends on which particular model is applied.

Ashurst [7] proposed a completely different approach, based on the baroclinic torque³. Azimuthal vorticity is produced by the density gradient over the flame front and the radial pressure gradient in the unburned swirling mixture. This induces an axial velocity towards the unburned mixture leading to higher flame speeds with decreasing density ratio for $\sigma > 3$.

Even though there are some similarities between the flashback process in the present study and the rapid flame propagation in vortex tubes, important differences exist, which limit the application of the theories developed so far. In our case, flashback occurs

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Amsterdam, The Netherlands, June 3–6, 2002; Paper No. 2002-GT-30075. Manuscript received by IGTI, Dec. 2001, final revision, Mar. 2002. Associate Editor: E. Benvenuti.

¹vortex bursting mechanism.

²back-pressure driven flame propagation mechanism.

³baroclinic push.

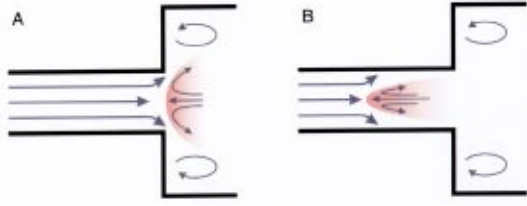


Fig. 1 Schematic sketch of a stable flame (a) and a flashback due to combustion induced vortex breakdown (b)

in a complex swirling highly turbulent flow. Starting from a position of the vortex breakdown at the burner exit, the flashback is closely linked to the sudden upstream propagation of the recirculation zone. The prerequisite for the upstream propagation of such a combustion induced vortex breakdown in the mixing tube is that the thermal and fluid dynamic properties inside the vortex core are favorable for a stable reaction during the flame propagation upstream.

It is known that high stretch rates or heat losses in the reaction zone can cause extinction, i.e., quenching of the flame, which in turn may lead to the collapse of the CIVB with a subsequent recovery of the steady flow field in the mixing tube. It will be shown later that the flame quenching is of great importance for the interpretation of the flashback limits of the investigated swirl burner. For this reason, analytical quenching theories are now briefly reviewed:

Several basic approaches for flame quenching have been proposed, [8–16], for diffusion as well as premixed flames in laminar and turbulent flows. In any case, a critical amount of energy has to be released while the unburnt gas mixture is in contact with the hot combustion products in order to maintain the chemical reaction. With increasing stretch rates, the heat release becomes increasingly limited by the reaction kinetics. As a consequence, the balance between the generation of products and the removal from the reaction zone due to turbulent mixing becomes more and more disturbed until the chemical reaction finally breaks down. Based on the laminar flame speed, characteristic length scales, and velocities, characteristic time scales for the flow and the reaction can be defined. If for the characterization of this balance the ratio of a characteristic chemical time τ_c and a characteristic flow or mixing time τ_u is used, flame quenching occurs at a critical value for the quench parameter C_{quench} defined as the ratio of both time scales:

$$\frac{\tau_c}{\tau_u} \geq C_{\text{quench}} \quad (2)$$

For quenching due to turbulent stretch Abdel-Gayed and Bradley et al. [8–11] proposed the characteristic flow time $\tau_{\lambda_T} = \lambda_T / u'$ based on the Taylor micro scale λ_T and the velocity fluctuations u' of the turbulent flow. Combined with $\tau_c = \delta_L / S_L$, where δ_L is the laminar flame thickness and S_L is the laminar flame speed, the Karlovitz stretch factor K is derived as

$$K = \frac{\tau_c}{\tau_{\lambda_T}} = \frac{\delta_L}{S_L} \cdot \frac{u'}{\lambda_T} = \frac{a}{S_L^2} \cdot \frac{u'}{\lambda_T} \quad (3)$$

where the laminar flame thickness is approximated by the thermal diffusivity a and the laminar flame speed S_L . It was shown in the past that hydrocarbon-air-mixtures can be quenched easier than hydrogen-air-mixtures because of the influence of the mass diffusivity D . Therefore, the Lewis number $Le = a/D$ is often considered to extend the quenching limits to fuels with high diffusivity. Based on experimental results a critical value of

$$K \cdot Le \geq 1.5 \quad (4)$$

was found, [10].

With $\bar{u}_a$ as the characteristic velocity and D_i as the characteristic length scale and assuming a perfectly stirred reaction zone, the characteristic flow time $\tau_u = D_i / \bar{u}_a$ can be interpreted as the residence time of the gases in the recirculation zone, [17]. Depending on the burner configuration, Eq. (2) can be transformed to

$$\frac{a}{S_L^2} \cdot \frac{\bar{u}_a}{D_i} \geq C_{\text{quench}} \quad (5)$$

For the lean blowout in gas turbine combustors, a similar approach leads to the Peclet number model, [17–20]. The stability limits of recirculating flames are described by the Peclet numbers based upon the characteristic length D_i , the thermal diffusivity a , and the blowoff velocity $\bar{u}_a$ or the laminar flame speed S_L , respectively,

$$Pe_u \sim Pe_{S_L}^2 \quad \text{with} \quad Pe_u = \frac{\bar{u}_a D_i}{a} \quad \text{and} \quad Pe_{S_L} = \frac{S_L D_i}{a} \quad (6)$$

Compared to Eq. (5), Eq. (6) can be written as

$$\frac{\bar{u}_a D_i}{a} \sim \frac{S_L^2 D_i^2}{a^2} \quad \text{or} \quad \frac{\bar{u}_a D_i}{a} = C_{\text{quench}} \cdot \frac{S_L^2 D_i^2}{a^2} \quad (7)$$

Interestingly, a quite similar Peclet number approach based on the critical gradient theory of Lewis and von Elbe [21] is also applied to describe the flashback in boundary layers, [12,14,15], although the physical background is completely different from the flashback process in the present study.

Experimental Setup

Figure 2 shows the schematic of the 200 kW atmospheric test rig. The fuel is added to the preheated air in a multistage mixing section and a homogeneous mixture is generated upstream of the swirler. The burner consists of a swirl generator and a cylindrical quartz glass mixing tube without centerbody, [1]. The flow profiles generated by the burner are tailored to provide a stable recirculation bubble in the combustion chamber for aerodynamic stabilization of the premixed flame, Fig. 3. As already mentioned, the circumferential velocity distribution is of major importance for the flame propagation in the swirling flow in the mixing tube. It is remarkable that flashback on the axis can occur despite the strong axial jet in the vortex core found in cold flow as well as in flow with stable reaction in the combustor.

Starting from stable operating conditions, the upstream flame propagation from the combustion chamber into the mixing tube (flashback) could be initiated by a continuous decrease of the air excess ratio λ at a constant air mass flow. As described earlier, [1], a flashback is not one continuous upstream propagation of the flame front, but rather consists of numerous shifts of the flame from the combustion chamber into the mixing tube and back

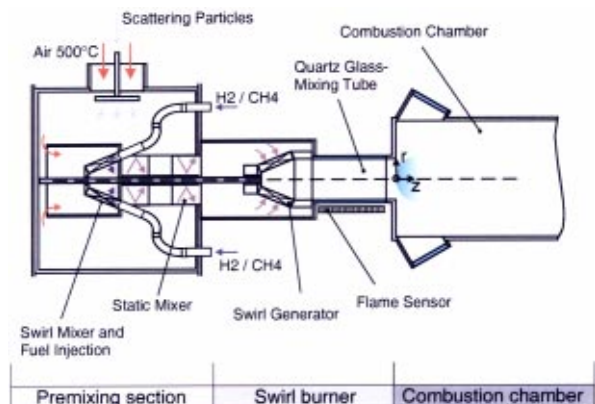


Fig. 2 Experimental facility

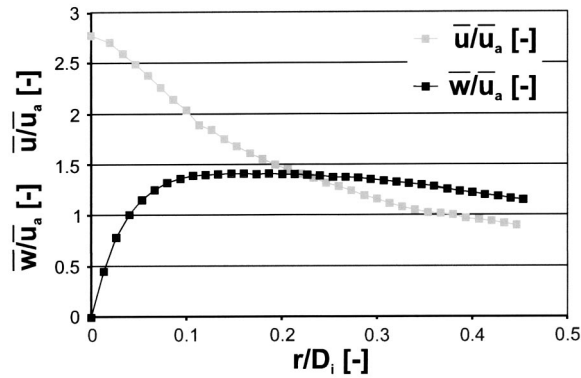


Fig. 3 Dimensionless axial ($\bar{u}/\bar{u}_a$) and circumferential ($\bar{w}/\bar{u}_a$) velocity profiles in the mixing tube

again. As long as the flow is able to recover in this way, the flame is not able to penetrate the entire mixing tube, but once the flame reaches the swirl generator, it stabilizes there due to high turbulence and low velocity regions in the vicinity of the swirl vanes trailing edges. As a consequence, the flame is no longer shifted downstream and the burner can be damaged once the stable reaction in the burner is established. For the sake of a precise definition of the flashback limit, it is defined in this study as the operating point at which the flame propagates through the entire mixing tube.

A special optical flame sensor was applied to detect the instantaneous flame position with high temporal and spatial resolution: A series of photodiodes is arranged along the mixing tube. Each individual diode generates a trigger signal in case of flame detection. In the presented study the signal from diode which is located closest to the swirl generator was used as the indication of flashback.

Fuels used in this study were natural gas and mixtures of natural gas and hydrogen. The natural gas consisted of 98.5 vol.% CH₄, 0.83 vol.% N₂, 0.08 vol.% CO₂ and 0.59 vol.% other hydrocarbons. Due to the high methane volume fraction, the properties of the natural gas are well represented by methane. Besides natural gas, hydrogen-methane-mixtures with the following mass fractions Y_{Fuel} and the corresponding volume fractions X_{Fuel} have been tested:

Mixture	$Y_{\text{CH}_4, \text{Fuel}}$	$Y_{\text{H}_2, \text{Fuel}}$	$X_{\text{CH}_4, \text{Fuel}}$	$X_{\text{H}_2, \text{Fuel}}$
1	1.00	0.00	1.000	0.000
2	0.95	0.05	0.7047	0.2953
3	0.90	0.10	0.5306	0.4694
4	0.85	0.15	0.4158	0.5842

Laminar Burning Velocities. Among the flow profiles and the density ratio, the laminar flame speed is expected to be one of the most important parameters governing flashback. In the tests, the use of hydrogen in the fuel allowed a wide variation of the flame speed. Unfortunately, the burning velocities of methane-hydrogen-mixtures are not well known for elevated initial temperature and very lean mixtures ($\lambda > 2.0$), [22–24]. Unlike the data for pure methane, the measured and computed flame speeds for very lean hydrogen-air-flames at ambient temperature vary in a wide range, [25–28], (Fig. 4). The database on flame speeds of mixtures becomes even weaker for initial temperatures from 100°C to 450°C, the range required for the analysis of the present experiments.

A correlation for methane-hydrogen-air-flames was proposed by Liu et al. [29]. The maximum flame speed as well as the flame

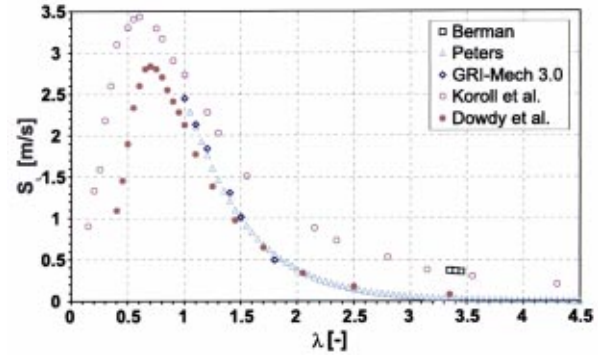


Fig. 4 Correlations, measurements, and computations of the laminar flame speed of H₂-air-flames at ambient temperature, $p=1$ bar

speed at stoichiometric conditions for the hydrogen-methane-mixtures can be calculated from those of the pure fuels:

$$\frac{1}{S_L} = \frac{X_{\text{CH}_4}}{S_{L, \text{CH}_4}} + \frac{X_{\text{H}_2}}{S_{L, \text{H}_2}} \quad (8)$$

A good agreement with the experimental results has been found by Liu for volume fractions of hydrogen in fuel below $X_{\text{H}_2, \text{Fuel}} < 0.70$. This value exceeds the maximum hydrogen content of $X_{\text{H}_2, \text{Fuel}} < 0.59$ used in the present study.

An application of this correlation to different equivalence ratios is shown in Figs. 5–8 (solid lines). The flame speeds for methane S_{L, CH_4} and hydrogen S_{L, H_2} are calculated according to Peters [26], who proposed correlations for both methane and hydrogen-air-flames for a wide range of initial temperatures up to 527°C:

$$S_L = F Y_{\text{Fuel}}^m e^{-G/T^0} \left(\frac{T_{\text{ad}} - T^0}{T_{\text{ad}} - T_{\text{mix}}} \right)^n \quad (9)$$

with

$$T^0 = \frac{E}{\ln \frac{B}{p}} \quad \text{and} \quad T_{\text{ad}} = a T_{\text{mix}} + b + c \phi + d \phi^2 + e \phi^3 \quad (10)$$

and the parameters for methane and hydrogen:

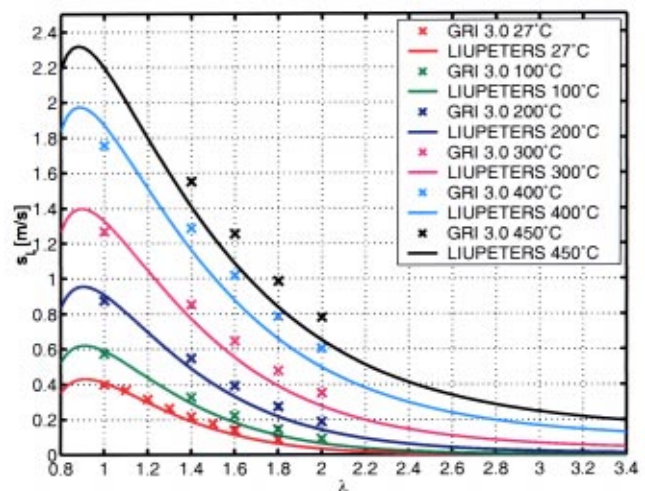


Fig. 5 Laminar flame speeds of CH₄-air-flames at various initial temperatures (mixture 1, 1 bar)

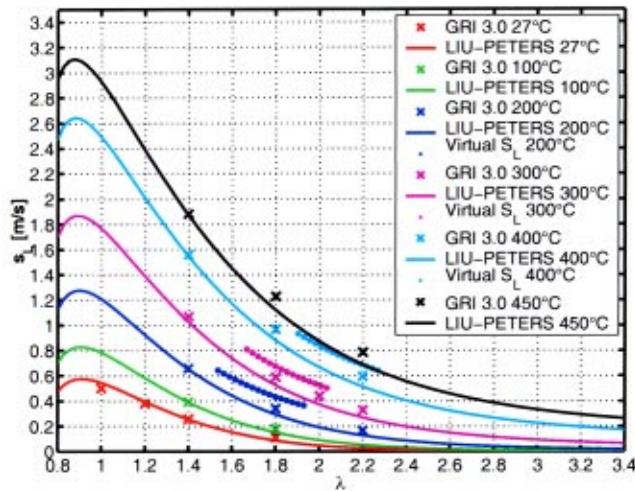


Fig. 6 Laminar flame speeds of $\text{CH}_4\text{-H}_2$ -mixtures in air (mixture 2, $Y_{\text{H}_2,\text{Fuel}}=0.05$, $p=1$ bar)

Fuel	a	b [K]	c [K]	d [K]	e [K]	
CH ₄	0.627	1270.15	−2449	6776	−3556	
	B [bar]	E [K]	m	F [cm/s]	G [K]	n
CH ₄	3.1557·10 ⁸	23873	0.565	22.176	−6444.27	2.516
	a	b [K]	c [K]	d [K]	e [K]	
H ₂	0.522	673.8	807.9	2515.6	−1765.9	
	B [bar]	E [K]	m	F [cm/s]	G [K]	n
H ₂	30044.1	10200.9	1.08721	1292880	2057.56	3.535

Some numerical calculations using a one-dimensional flame code, [30], and the GRI-Mech 3.0, [31], mechanism have been made to verify the laminar flame speeds of Eq. (8) (crosses in Figs. 5 to 8). With the exception of small deviations for elevated temperatures and methane-air-flames (lower values from the correlation than those calculated with detailed chemistry) both methods coincide very well. One the other hand, it has to be taken into

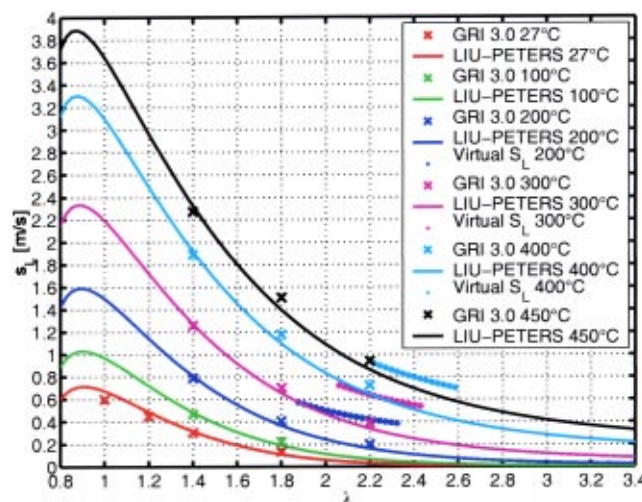


Fig. 7 Laminar flame speeds of $\text{CH}_4\text{-H}_2$ -mixtures in air (mixture 3, $Y_{\text{H}_2,\text{Fuel}}=0.10$, $p=1$ bar)

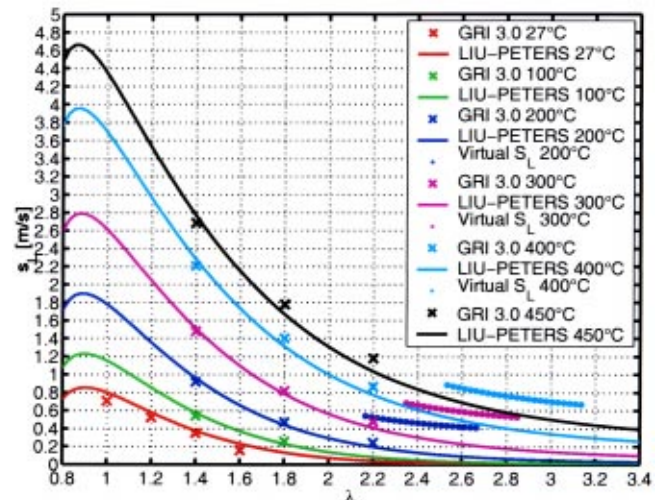


Fig. 8 Laminar flame speeds of $\text{CH}_4\text{-H}_2$ -mixtures in air (mixture 4, $Y_{\text{H}_2,\text{Fuel}}=0.15$, $p=1$ bar)

consideration that the predicted flame speeds for lean hydrogen-air-flames are considerably lower than many experimental values (e.g., Koroll et al. [25] and Berman [27], see Fig. 4 for comparison). Because of this deviation it must be concluded that the current knowledge concerning flame speeds in very lean mixtures of methane and hydrogen ($\lambda > 1.5$) is limited and that the theoretical methods seem to underpredict many of the experimentally found values, [32]. It is shown later that these big relative errors of the evaluated laminar flame speed S_L cause accuracy problems in applying Peclet number models.

Experimental Results

The air excess ratios λ at the flashback limits as a function of the mean flow rate are plotted in Figs. 9–12. In the experiments, the air mass flow has been varied from 60 g/s to 160 g/s at pre-mixing temperatures from 100°C to 450°C for mixture 1 and from 200°C to 400°C for mixture 2, 3, and 4, respectively. In order to examine the reproducibility, the experiments have been repeated up to five times for each operating point.

Obviously, the scattering of the experimental results increases slightly with decreasing preheating temperature T_{mix} . Low mean flow rates represent another limit of the experimental procedure applied. As the thermal power and therefore the emitted light of the reaction zone decreases, the optical detection of the flashback becomes increasingly difficult, particularly in case of the weakly radiating hydrogen flames. Nevertheless, the flashback limits are remarkably reproducible in the light of the high complexity of the

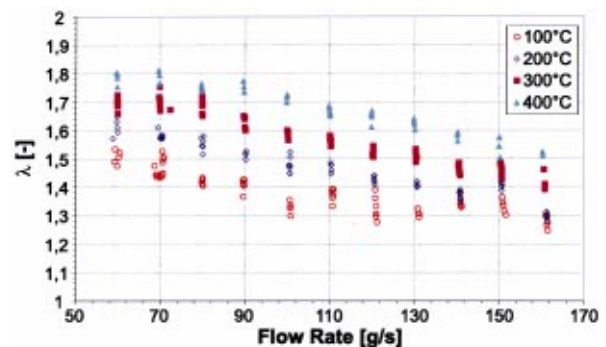


Fig. 9 Air excess ratio λ at the flashback limit (mixture 1, $Y_{\text{H}_2,\text{Fuel}}=0$)

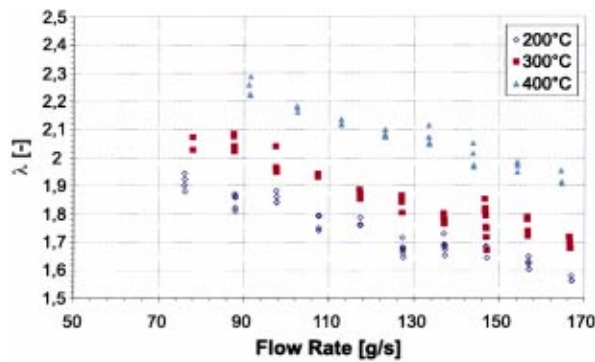


Fig. 10 Air excess ratio λ at the flashback limit (mixture 2, $Y_{H_2, fuel}=0.05$)

process. For all fuel mixtures an almost linearly declining correlation between λ and the flow rate can be found. Higher preheating temperatures or lower flow rates promote flashback. With increasing hydrogen fraction, the flashback occurs at extremely high air excess ratios λ . Mixture 4 produces flashback even for low preheating temperatures at $\lambda \geq 2$.

Figures 9 to 12 show the influence of the air mass flow on the flashback limits. The intuitive assumption that higher mass flows require lower air excess ratios λ is qualitatively supported, but this effect is very small. At a fixed temperature of, e.g., 200°C, a substantial step in the mean flow rate from 60 g/s to 160 g/s leads to a moderate decrease of λ from 1.6 to 1.3 (Fig. 9). Concerning its technological dimension, this is a very interesting finding, since flashback safety cannot be considerably improved by increasing the flow velocity as often expected, if the CIVB is the cause.

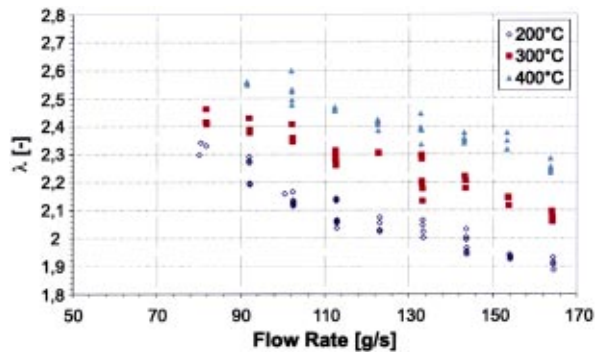


Fig. 11 Air excess ratio λ at the flashback limit (mixture 3, $Y_{H_2, fuel}=0.10$)

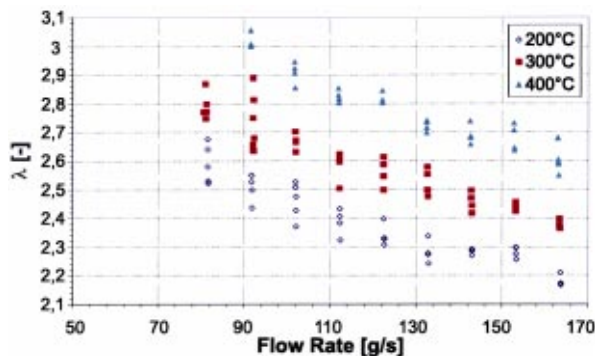


Fig. 12 Air excess ratio λ at the flashback limit (mixture 4, $Y_{H_2, fuel}=0.15$)

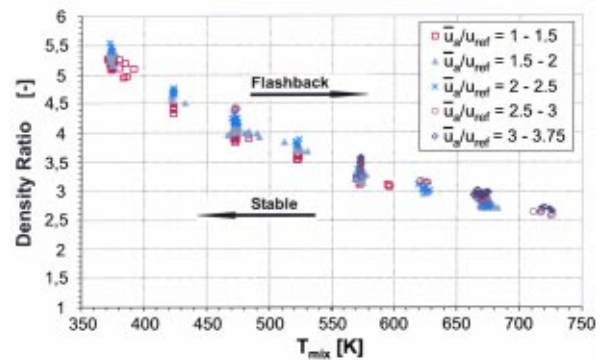


Fig. 13 Density ratio versus premixing temperature at the flashback limit, different mean velocities (mixture 1)

Since only one burner configuration was tested, the mean axial velocity, the circumferential velocity and the velocity fluctuations are proportional to each other. An increase of the mass flow as well as the preheating temperature raise the mean axial as well as the circumferential velocity, which according to equation 1 increases the driving force for the upstream flame propagation. For this reason, it was not possible to determine the quantitative influence of this ratio on the flashback limits with the present experimental approach. However, the theoretical finding that the flame propagation velocity is closely linked to the circumferential velocity provides one opportunity to explain the fairly low influence of the flow speed. Some of the mentioned theories lead to the final conclusion that swirling flows can be divided into two classes, one which is prone to flashback independently from the mass flow and another which is flashback safe for all mass flows. In the light of this fact, the low sensitivity of the flashback limit concerning the axial flow velocity becomes plausible. This sensitivity is much lower than in the case of a flashback in the boundary layer because the latter mechanism depends on the velocity gradient at the wall, which is a function of the mass mean axial velocity.

As shown in Figs. 13–16, the density ratios at the flashback limit vary for different premixing temperatures at constant velocities. This is a very important finding, because it proves that vortex dynamics is not the only relevant governing effect, as found in generic theoretical studies. It is obvious that the acceleration of the chemical kinetics at higher reactant temperatures supports the flashback process. An attempt to find a linear correlation on the basis of the laminar flame speed failed. This is not a surprise, since the experiments reveal the highly turbulent nature of the process. It is then straightforward to consider the turbulent flame speed S_T for the investigated strongly turbulent flows instead.

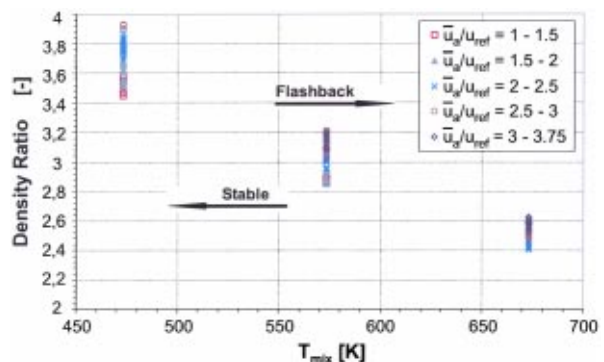


Fig. 14 Density ratio versus premixing temperature at the flashback limit, different mean velocities (mixture 2)

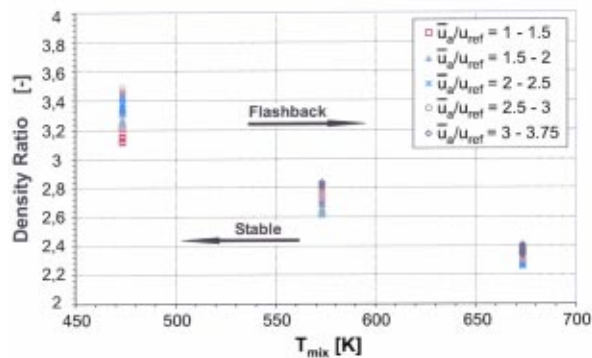


Fig. 15 Density ratio versus premixing temperature at the flashback limit, different mean velocities (mixture 3)

However, S_T does not depend considerably on the laminar flame speed for the turbulence level found in the mixing tube, [1], using a very simple law $S_T \approx S_L + u'$. In the present case, the velocity fluctuations u' exceed the laminar flame speed by one order of magnitude and the turbulent flame speed can be approximated by $S_T \approx u'$. Assuming that u' is proportional to the mean burner velocity, also the turbulent burning velocities follow this proportionality and the unrealistic classification of swirling flows into inherently flashback prone and flashback safe cases, respectively, is obtained again. Moreover, neither the effect of the reactant temperature nor fuel effects (Figs. 13 to 16) can be explained with this approach.

It was shown experimentally, [1], that the recirculation zone of the CIVB can be reasonably approximated by a perfectly stirred reactor, since the degree of mixing is comparable to the turbulent mixing in the recirculation zone of premixed swirl stabilized flames. Due to the interaction of the reaction zone with the surrounding unburned gas mixture, the entrainment of the cold fuel-air-mixture into the CIVB leads to a cooling of the reacting gases. If the heat release of the reaction cannot balance the heat losses, the flame is quenched, the CIVB becomes unstable and a further upstream flame propagation is suppressed. The nature of quenching was already reported before: Measurements of the laser-induced fluorescence of the OH-radical in the CIVB revealed a highly disrupted flame structure as the flame propagation came to rest in the mixing tube, [1].

Since there is some evidence of quenching as the governing effect, the quench factor C_{quench} at the flashback limit was calculated for all experiments according to Eq. (5). Although there is some scatter in the data and a weak effect of the density ratio may exist, an almost constant value for C_{quench} for all experiments is found, as shown in Fig. 17. This supports the hypothesis that

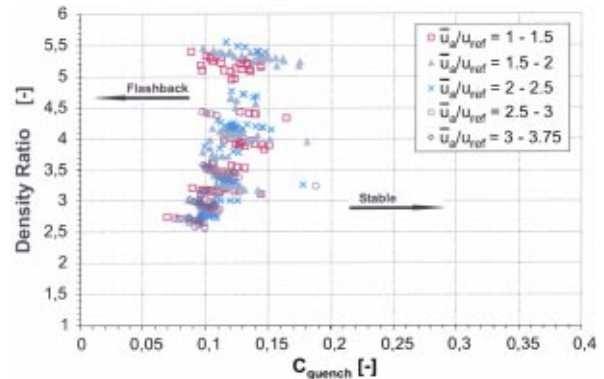


Fig. 17 Quenching constant C_{quench} at the flashback limit (mixture 1)

although the density ratio certainly influences the upstream flame propagation obviously the quenching is the dominating criterion.

This finding has an important technological aspect, since, once the quench factor C_{quench} has been determined experimentally, the flashback limits can be calculated for other operating conditions, whereby the observed data scattering of about $\pm 30\%$ (Fig. 17) corresponds to a variation of the air excess ratio λ of only $\pm 5\%$. Moreover, the quench factor C_{quench} represents a quality measure for specific burner designs.

The quench factor C_{quench} can be easily translated into a Peclet number model for flame quenching. If the experiments are correlated in terms of the Peclet numbers Pe_u and Pe_{S_L} (see Eqs. (6), (7)) the results shown in Fig. 18 are obtained. Although the flashback process is of complex nature and shows some scatter in the experiments, a very good correlation is obtained, which supports again the hypothesis of flame quenching. Interestingly, the behavior of the CIVB can be formally accessed with the same parameters as the lean blowout limit of swirl stabilized premixed flames.

For all of four investigated fuel mixtures a similar behavior was found. However, the evaluation of the quenching constant C_{quench} was not as successful for the methane-hydrogen-mixtures (mixtures 2, 3, and 4) as for methane-air-flames. As $C_{\text{quench}} \sim 1/S_L^2$ strongly depends on the laminar flame speed, the analysis of the quenching limit suffers if the laminar flame speeds in very lean hydrogen-methane-air flames cannot be determined with sufficient precision, a problem which was already mentioned. Assuming the same quenching constants C_{quench} as determined for methane before, it can be indirectly checked, which laminar flame speeds are needed in order to make the model fully consistent also for the natural gas-hydrogen mixtures. $C_{\text{quench}} = 0.12$ presumed, virtual

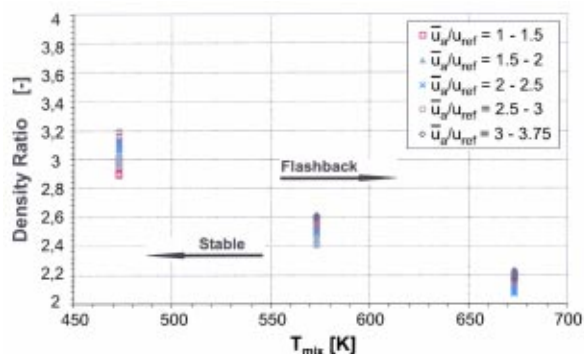


Fig. 16 Density ratio versus premixing temperature at the flashback limit, different mean velocities (mixture 4)

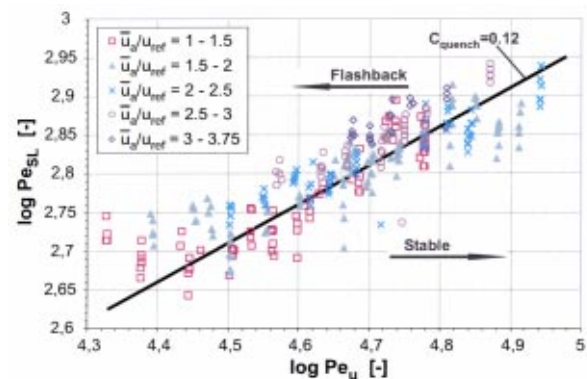


Fig. 18 Correlation of flame quenching at the flashback limits based on Peclet numbers (mixture 1)

laminar flame speeds for methane-hydrogen mixtures can be calculated ($S_L = \sqrt{au_a/0.12D_i}$). These virtual laminar flame speeds are marked as dots in Figs. 6–8.

With increasing hydrogen fraction and air excess ratio λ , a deviation between virtual and calculated flame speeds is clearly visible, but if the huge relative scatter of the measurements in the very lean regime of hydrogen-air-flames (Fig. 4) is taken into account, the deviation is tolerable. Additional work on burning velocities, or more precise on the characteristic chemical time τ_c of these flames is required to clarify the remaining uncertainties. For this reason perfectly stirred reactor studies are planned to evaluate τ_c directly. In evaluating the data from the hydrogen containing fuels, Lewis number effects on quenching could not be considered so far. This will be an area of future interest.

Conclusions

- The combustion-induced vortex breakdown CIVB is the governing mechanism of the upstream propagation of the vortex core in the investigated swirl burner.
- An increase of the flow velocity in the burner leads to only moderate improvements of the CIVB flashback safety.
- Quenching of the reaction in the vortex core is the most dominant process regarding the flashback limit of the investigated burner configuration. As long as the flame in the vortex core extinguishes due to the interaction of the unburnt gases and the reaction zone, a flashback through the entire mixing tube can be prevented.
- A quench factor was derived. When applied to a particular burner design, it characterizes the CIVB flashback resistance.
- Using this quench factor, CIVB flashback limits for other than the measured operating points can be derived and the flashback behavior of a burner can be characterized without need for an excessive number of experiments.
- A Peclet number model has been applied successfully to correlate the flashback limits of the investigated burner configuration.
- The applicability of the presented models has to be reviewed for different burner configurations in the future. Especially modifications in the vortex core size are expected to influence the flashback limits considerably.

Acknowledgment

It is gratefully acknowledged that the experimental work on CIVB presented in this paper has been supported by ALSTOM-Power (CH).

Nomenclature

a	= thermal diffusivity [m ² /s]
C_{quench}	= quenching constant
D	= diffusion coefficient [m ² /s]
D_i	= inner mixing tube diameter, burner diameter [m]
K	= Karlovitz stretch factor
Le	= Lewis number
p	= pressure [bar]
Pe_{S_L}	= Peclet number based on the laminar flame speed
Pe_u	= Peclet number based on the mean burner velocity
r	= radial coordinate [m]
S_f	= flame speed [m/s]
S_L	= laminar burning velocity [m/s]
S_T	= turbulent burning velocity [m/s]
T_{ad}	= adiabatic flame temperature [°C]
T_{mix}	= fuel/air temperature in the mixing tube [°C]
u	= axial velocity [m/s]
u_{ref}	= reference axial velocity [m/s]
u'	= r.m.s. of fluctuating axial velocity [m/s]
u_a	= mass mean axial velocity in the mixing tube [m/s]
w	= circumferential velocity [m/s]
X_{Fuel}	= volume fraction in fuel

Y_{Fuel}	= mass fraction in fuel
δ_L	= laminar flame thickness [m]
λ	= air excess ratio (1/ ϕ)
λ_T	= Taylor micro scale [m]
ϕ	= equivalence ratio
ρ_b	= density of burnt gas [kg/m ³]
ρ_u	= density of unburnt gas [kg/m ³]
σ	= density ratio across the flame front
τ_c	= characteristic chemical time [s]
τ_u	= characteristic flow time, residence time [s]
τ_{λ_T}	= characteristic flow time based on the Taylor micro scale [s]

References

- [1] Fritz, J., Kröner, M., and Sattelmayer, T., 2001, "Flashback in a Swirl Burner With Cylindrical Premixing Zone," ASME Paper No. 2001-GT-0054.
- [2] McCormack, P. D., Scheller, K., Mueller, G., and Tisher, R., 1972, "Flame Propagation in a Vortex Core," *Combust. Flame*, **19**, pp. 297–303.
- [3] Asato, K., Wada, H., Hiruma, T., and Takeuchi, Y., 1997, "Characteristics of Flame Propagation in a Vortex Core: Validity of a Model for Flame Propagation," *Combust. Flame*, **110**, pp. 418–428.
- [4] Chomiak, J., 1997, "Dissipation Fluctuations and the Structure and Propagation of Turbulent Flames in Premixed Gases at High Reynolds Numbers," *17th International Symposium on Combustion*, The Combustion Institute, Pittsburgh, PA, pp. 1665–1673.
- [5] Umemura, A., and Tomita, K., 2001, "Rapid Flame Propagation in a Vortex Tube in Perspective of Vortex Breakdown Phenomena," *Combust. Flame*, **125**, pp. 820–838.
- [6] Ishizuka, S., Hamasaki, T., Koumura, K., and Hasegawa, R., 1998, "Measurements of Flame Speeds in Combustible Vortex Rings: Validity of the Back-Pressure Drive Flame Propagation Mechanism," *27th International Symposium on Combustion*, The Combustion Institute, Pittsburgh, PA, pp. 727–734.
- [7] Ashurst, W. M. T., 1996, "Flame Propagation Along a Vortex: The Baroclinic Push," *Combust. Sci. Technol.*, **112**, pp. 175–185.
- [8] Abdel-Gayed, R. G., Bradley, D., Hamid, M. N., and Lawes, M., 1984, "Lewis Number Effects on Turbulent Burning Velocity," *20th Symposium (International) on Combustion*, The Combustion Institute, Pittsburgh, PA, pp. 505–512.
- [9] Abdel-Gayed, R. G., and Bradley, D., 1985, "Criteria for Turbulent Propagation Limits of Premixed Flames," *Combust. Flame*, **62**, pp. 61–68.
- [10] Abdel-Gayed, R. G., Bradley, D., and Lung, F. K. K., 1989, "Combustion Regimes and the Straining of Turbulent Premixed Flames, Short Communications," *Combust. Flame*, **76**, pp. 213–218.
- [11] Bradley, D., 1992, "How Fast Can We Burn?" *24th Symposium (International) on Combustion*, The Combustion Institute, Pittsburgh, PA, pp. 247–262.
- [12] Putnam, A. A., and Jensen, R. A., 1948, "Application of Dimensionless Numbers to Flash-Back and Other Combustion Phenomena," *3rd International Symposium on Combustion, Flame and Explosion Phenomena*, The Combustion Institute, Pittsburgh, PA, pp. 89–98.
- [13] Kremer, H., Minx, E., and Rawe, R., 1976, "Flammenstabilität bei Gasbrennern mit Gebläse," *Forschungsbericht des Landes Nordrhein-Westfalen*, Nr. 2567/Fachgruppe Energie.
- [14] Prade, B., and Lenze, B., 1992, "Experimental Investigation in Extinction of Turbulent Non-Premixed Disk Stabilized Flames," *24th Symposium (International) on Combustion*, The Combustion Institute, Pittsburgh, PA, pp. 369–375.
- [15] Turns, S. R., 2000, *An Introduction to Combustion*. 2nd Ed., Gordon and Breach, New York.
- [16] Chomiak, J., 1990, *Combustion—A Study in Theory, Fact and Application*.
- [17] Hoffmann, S., 1994, "Untersuchungen des Stabilisierungsverhaltens und der Stabilitätsgrenzen von Drallflammen mit innerer Rückströmzone," Ph.D. thesis, TH Karlsruhe.
- [18] Hillemanns, R., 1988, "Das Strömungs- und Reaktionsfeld sowie Stabilisierungseigenschaften von Drallflammen unter dem Einfluss der inneren Rezirkulationszone," Ph.D. thesis, TH Karlsruhe.
- [19] Hoffmann, S., Habisreuther, P., and Lenze, B., 1994, "Development and Assessment of Correlations for Predicting Stability Limits of Swirling Flames," *Chem. Process Eng. (London)*, **33**, pp. 393–400.
- [20] Hoffmann, S., Lenze, B., and Eickhoff, H., 1998, "Results of Experiments and Models for Predicting Stability Limits of Turbulent Swirling Flames," *ASME J. Eng. Gas Turbines Power*, **120**, pp. 311–316.
- [21] Lewis, B., and von Elbe, G., 1987, *Combustion Flames and Explosion of Gases*, Third Ed.
- [22] Nastoll, W., 1989, "Untersuchung zur instationären turbulenten Flammenausbreitung in geschlossenen Behältern," Ph.D. thesis, TH Karlsruhe.
- [23] Scholte, T. G., and Vaags, P. B., 1959, "Burning Velocities of Mixtures of Hydrogen, Carbon Monoxide and Methane With Air," *Combust. Flame*, **3**, pp. 511–524.
- [24] Yu, G., Law, C. K., and Wu, C. K., 1986, "Laminar Flame Speeds of Hydrocarbon+Air Mixtures With Hydrogen Addition," *Combust. Flame*, **63**, pp. 339–347.
- [25] Koroll, G. W., Kumar, R. K., and Bowles, E. M., 1993, "Burning Velocities of Hydrogen-Air Mixtures," *Combust. Flame*, **94**, pp. 330–340.

- [26] Peters, N., 1994, "Turbulente Brenngeschwindigkeit," *Abschlussbericht zum Forschungsvorhaben Pe 241/9-2*.
- [27] Berman, M., 1984, Sandia Laboratories Report, SAND84-0689.
- [28] Dowdy, D. R., Smith, D. B., Taylor, S. C., and Williams, A., 1990, "The Use of Expanding Spherical Flames to Determine Burning Velocities and Stretch Effects in Hydrogen/Air Mixtures," *Combust. Flame*, pp. 325–332.
- [29] Liu, Y., 1991, "Untersuchung zur Stationären Ausbreitung turbulenter Vormischflammen," Ph.D. thesis, TH Karlsruhe.
- [30] Kee, R. J., Grcar, J. F., Smooke, M. D., Miller, J. A., and Meeks, E., 1992, "A Fortran Program for Modeling Steady Laminar One-Dimensional Premixed Flame," *Sandia National Laboratories Report SAND85-8240*.
- [31] Smith, G. P., Golden, D. M., Frenklach, M., Moriarty, N. W., Eiteneer, B., Goldenberg, M., Bowman, C. T., Hanson, R. K., Song, S., Gardiner, W. C., Jr., Lissianski, V. V., and Qin, Z., GRI-Mech 3.0. http://www.me.berkeley.edu/gri_mech/.
- [32] Moser, V., 1997, "Simulation der Explosion Magerer Wasserstoff-Luft-Gemische in Großskaligen Geometrien," Ph.D. thesis, RWTH Aachen.

Impact of Ethane and Propane Variation in Natural Gas on the Performance of a Model Gas Turbine Combustor

R. M. Flores

V. G. McDonell

G. S. Samuelsen

UCI Combustion Laboratory,
University of California,
Irvine, CA 92697-3550

In the area of stationary power generation, there exists a growing interest in understanding the role that gaseous fuel composition plays on the performance of natural gas-fired gas turbine systems. In this study, an atmospherically fired model gas turbine combustor with a fuel flexible fuel/air premixer is employed to investigate the impact of significant amounts of ethane and propane addition into a baseline natural gas fuel supply. The impacts of these various fuel compositions, in terms of the emissions of NO_x and CO, and the coupled impact of the degree of fuel/air mixing, are captured explicitly for the present system by means of a statistically oriented testing methodology. These explicit expressions are also compared to emissions maps that encompass and expand beyond the statistically based test matrix to verify the validity of the employed statistical approach.

[DOI: 10.1115/1.1584480]

Motivation

In an effort to meet increasingly stringent emissions regulations, a number of researchers and manufacturers are now using lean-premixed gas turbine combustion as an alternative to other more traditional modes of gas turbine operation. Due to the complex nature of combustion systems, however, there remains much that is unknown about operating a combustion system lean; and there is even less known about operating lean and premixed. One unknown that is of growing concern is the effect that gaseous fuel variability may have on the performance of a lean-premixed natural gas-fired gas turbine system (e.g., [1]).

Recent work by Flores et al. [2] revealed the degree to which the performance of a model gas turbine combustor was dependent upon the composition of the fuel. In that study, several different gaseous fuel blends were employed that included the addition of significant amounts of ethane or propane to the baseline fuel of natural gas. These fuels represented some of the extreme levels of ethane and/or propane that may be found within the natural gas supply of the United States (Liss et al. [3]).

The model combustor employed in Flores et al. [2] was operated in a rapid mix mode of operation, and subtle differences in the fuel distribution—associated with differing momentum ratios for the various fuels—were found to affect the performance of the system. To extend the findings to a fuel/air premixer more representative of advanced practical mixers, the current study was undertaken. In addition, the present study utilizes experimental methods specifically directed at providing an explicit, quantitative relationship between the fuel composition and operation of the model combustor.

Objectives

The objectives of the current study are to (1) perform a series of statistically based experiments examining the effects of fuel composition on the performance of a model gas turbine combustor, (2) to evaluate the effectiveness of the this statistical methodology with emissions and performance maps encompassing the model

space, and (3) develop an explicit relationship between emissions performance and the combustor inlet fuel distribution and composition.

Approach

The approach taken is to develop and apply a modular combustor premixer with a flexible fuel injection system to provide control over the placement and distribution of the fuel within the combustion air stream. The fuel injection system is based on radial fuel jets that are located along the stem of the centerbody and the surrounding annulus wall of the combustor premixer. The ability to control the mixing and blending of the air/fuel mixture comprises a significant variable, along with the composition of the gaseous fuel blend, in the execution of a statistically based series of experiments.

The statistical experimentation is followed with the generation of emissions maps, as was done previously for a different system, [2]. These maps are based upon the emissions of CO, NO_x , and the measured lean blowout (LBO) limits (ϕ_{LBO}) as measured for several specific fuel compositions, and will be generated as a function of radial fuel split and equivalence ratio. Additional insight concerning the effectiveness of this statistically based testing methodology is drawn from the measured distribution of the fuel, which is characterized at the inlet plane using an extractive probe.

Experiment

Test Facility. The test facility utilized provides a wide range of operating conditions and flow metering. The test stand is designed to operate at 1 atm with air inlet temperatures up to 800 K. The combustor premixer and quartz combustor liner, shown schematically in Fig. 1, are attached to a three-dimensional traverse which allows the system to be moved as necessary about fixed diagnostic equipment and/or probes to map out points both within and at the exit plane of the combustor.

Combustor Premixer. A cross-sectional schematic of the combustor premixer from Fig. 1 is provided in Fig. 2 depicting the flow path of the combustion air entering the combustor premixer. Figure 3 shows the various fuel injection circuits available for the system, as well as their relative locations. The fuel injection system employs a total of ten gaseous mass flow controllers for monitoring and controlling the amount and placement of the

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Amsterdam, The Netherlands, June 3–6, 2002; Paper No. 2002-GT-30080. Manuscript received by IGTI, December 2001, final revision, March 2002. Associate Editor: E. Benvenuti.

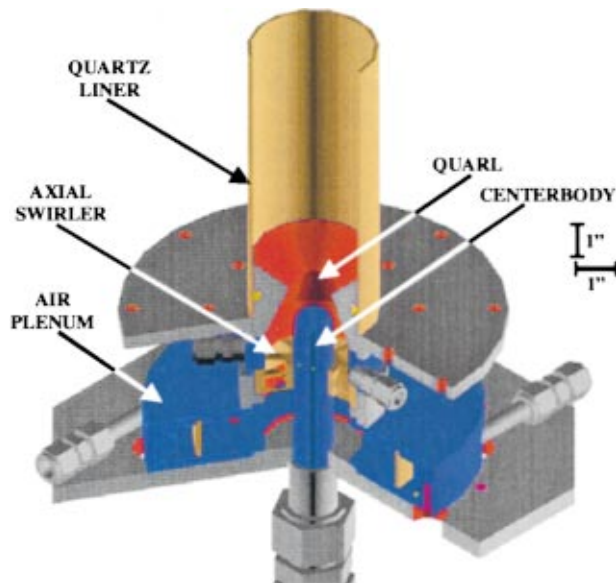


Fig. 1 Relative location of centerbody and combustion pre-mixer (contraction and exit sampling sections not shown)

blended fuel within the annulus of the combustor pre-mixer. This provides for a multipoint approach to fuel and air mixing with the ability to control fuel flow splits between the various fuel injection circuits.

With the present system, three independent fuel injection circuits are available. One option is to inject fuel radially from the centerbody into the surrounding, swirling air stream. This centerbody injection circuit, labeled “CB” in Fig. 3, consists of eight

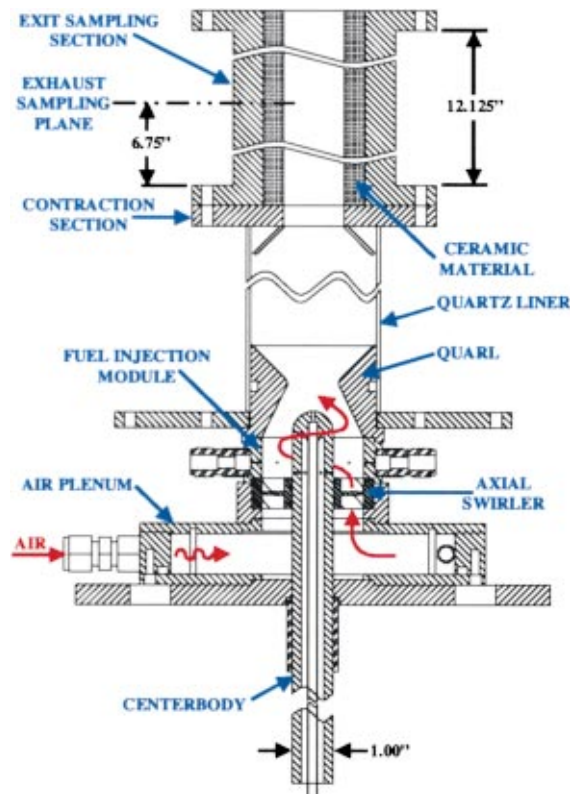


Fig. 2 Cross section and air flow path of the system with an axial swirler and exit sampling section

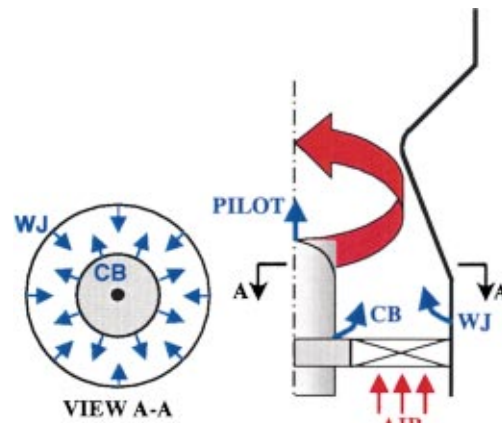


Fig. 3 Relative location and number of the fuel injection circuits

equally spaced fuel holes located circumferentially along the stem of the centerbody. The second fuel injection option is designed to inject the fuel radially from the surrounding annulus wall. This wall injection circuit, labeled “WJ” in Fig. 3, also consists of eight equally spaced fuel holes in positions that are staggered with respect to the centerbody injection holes. The final fuel injection option is to inject the fuel axially into the air stream. This axial injection circuit, labeled “PILOT” in Fig. 3, consists of a single fuel injection hole located at the tip of the centerbody.

By adjusting the distribution of the fuel between the radial and axial fuel injection circuits, the overall radial profile of the fuel distribution may be altered. This is demonstrated in Fig. 4, where it is illustrated that, by varying the fuel from the wall injection circuit to the centerbody injection circuit, the fuel distribution can be weighted to either the centerline of the combustor pre-mixer annulus or its wall. Likewise, additional fuel distribution weighting towards the centerline of the annulus may be achieved with the pilot injection circuit. Figure 5 presents representative radial fuel distribution profiles with and without pilot fuel for a given operating condition while illustrating that the composition of the fuel does not play a strong role in the overall curvature of the radial fuel distribution. This degree of fuel distribution control exceeds the level of control achieved previously in Flores et al. [2].

Additionally, the baseline configuration of the combustor pre-mixer utilizes an eight-vane, 45°-axial swirler followed by a fuel injection module (providing the fuel jets for the annulus wall), and

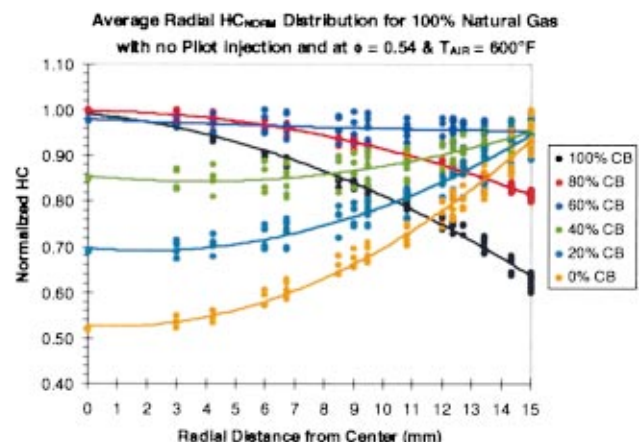


Fig. 4 Typical radial profile of fuel as a function of percent centerbody

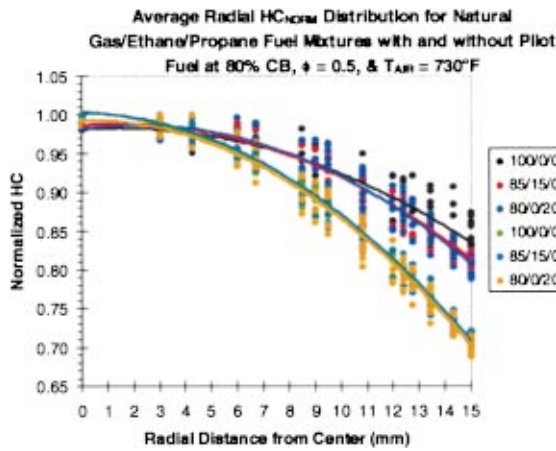


Fig. 5 Representative radial fuel distribution with and without pilot fuel

a quarl. The nominal firing rate for the system is 35 kW at 152.5 lbm/hr of air, though the fuel flow rates were varied to assess the performance of the system and applicability of the statistical results at different equivalence ratios. The air preheat temperature of the system was fixed at 660 K, which was achieved using an inline, nonvitiated, electrical air heater.

Diagnostics. Exhaust emissions of carbon monoxide (CO), carbon dioxide (CO₂), hydrocarbons (HC), oxygen (O₂), and nitrogen oxides (NO_x) were measured using Horiba Ltd. analyzers. These instruments are part of an integrated sampling and computer data acquisition system.

A 12.7-mm o.d. water-cooled, stainless steel bulk emissions probe is used to sample the exhaust emissions downstream of the exit plane of the combustor as shown in Fig. 1. This probe is designed to take an integrated average measurement of the emissions over the diameter of the sampling plane via five area-weighted sampling ports. The water in the probe is heated to 325 K to protect the probe and quench the sample while avoiding condensation of water vapor inside the probe. The emissions are pumped through a Teflon line heated to 408 K to prevent water condensation, and the sample is then split into two streams. The NO_x stream goes through a converter to reduce any NO₂ to NO prior to the water drop out.

All of the emissions measurements from the analyzers are recorded using a digital data acquisition system. The data measurements from the analyzers are sampled and averaged over a 60-second period at a rate of 4 Hz. Error analysis and repeatability studies conducted on the fuel/air system and emission measurement equipment established an uncertainty of ± 0.5 ppm (cor-

Table 2 Gas compositions utilized

Blend	Wobbe Index* MJ/m ³	S.G. Relative to Air	LHV MJ/m ³
natural gas	44.3	0.576	33.6
85% natural gas/ 15% ethane	46.8	0.645	37.6
80% natural gas/ 20% propane	50.5	0.765	44.2

*Wobbe Index = $LHV/[SG]^{1/2}$ (Meier et al. [9])

rected to 15% O₂ by volume), for both CO and NO_x. Distributions of nonreacting fuel were mapped with a 12.7 mm o.d. stainless steel probe and a high range flame ionization detector (Horiba Model FIA-236-1) at the throat of the quarl, immediately upstream of the quartz combustor liner. All analyzers used had an accuracy of $\pm 1\%$ full scale.

Fuel Blending. The fuel blending system previously described in Flores et al. [2] is also employed for the current study. As before, a subset of a comprehensive fuel blending system is utilized to blend streams of natural gas, ethane, and propane using a set of Brooks gaseous mass flow controllers. The blended gas stream pressurizes a fuel manifold and bypass line connected to a backpressure regulator. This fuel manifold supplies the necessary fuel at pressure to a second set of mass flow controllers that comprise the fuel injection circuits of the combustor premixer. The entire process of monitoring and controlling the mass flow controllers is performed using a LabView based control program that allows the fuel composition and flow splits to be specified as desired. The overall accuracy of the blended fuel flow rate is $\pm 2\%$.

For the present study, the baseline fuel used is natural gas. A typical constituent composition of the available natural gas is summarized below in Table 1. The major constituent of the natural gas is methane, which is responsible for over 96% of the fuel's volume. Since a focus of this study is to examine the impact of significant increases in ethane and propane, the impact of the remaining minor constituents should be rendered negligible in comparison.

For the present study, three fuel compositions were considered for the emissions and performance maps to be compared with the statistical experiments. These three compositions represent the extremes in the fuel composition as described in the motivation section. Some of the associated properties are also summarized in Table 2.

Results

The results of the statistical experimentation are first presented and discussed. This is followed with a summary of the emissions maps that were generated. These emissions maps are then compared with the results of the statistical methodology to ascertain the validity of the statistical approach and to determine if the trends detected hold true under other conditions not directly considered in the statistical test matrix. Finally, a discussion of the relevant NO_x formation pathways is provided to ascertain the source of the NO_x generated during the course of this study.

Design of Experiments. The statistical approach applied to this study is based on a Design of Experiments (DoE) methodology, the fundamental aspects of which may be found in Box, Hunter, and Hunter [4]. The DoE approach can be described as a systematic series of tests, in which purposeful changes are made to input factors of a process, so that one can observe and identify the reasons for the changes in the output responses. A benefit of this technique is its ability to avoid testing one factor at a time by incorporating randomization and multifactorial experiments. By randomizing the sequence of experiments to be conducted, statis-

Table 1 Typical natural gas composition

n	Constituent	MF y(i)%
1	methane	CH ₄ 96.975
2	ethane	C ₂ H ₆ 0.982
3	propane	C ₃ H ₈ 0.109
4	iso-butane	C ₄ H ₁₀ 0.014
5	n-butane	C ₄ H ₁₀ 0.015
6	iso-pentane	C ₅ H ₁₂ 0.004
7	n-pentane	C ₅ H ₁₂ 0.004
8	C ₆	C ₆ H ₁₄ 0.001
9	C ₇	C ₇ H ₁₆ 0.001
10	C ₈	C ₈ H ₁₈ 0.000
11	CO ₂	1.574
12	O ₂	0.000
13	N ₂	0.322
Totals		100.00

Table 3 Sample of randomized test matrix

A: % CH ₄	B: % C ₂ H ₆	C: % C ₃ H ₈	D: % Pilot	E: % CB
86.25	7.50	6.25	0.00	70
100.00	0.00	0.00	5.00	100
86.67	0.00	13.33	0.00	40
86.67	5.00	8.33	2.50	100
⋮	⋮	⋮	⋮	⋮
88.13	3.75	8.13	5.00	100
93.33	0.00	6.67	2.50	70
80.00	15.00	5.00	0.00	70

tical and probability tools and techniques may be employed in the analysis of the results, including quantifying the amount of variability and uncertainty resulting from uncontrollable input factors while simultaneously identifying those parameters that have the most impact on the output of the system.

In general, there are two types of experiments that may be conducted under the DoE approach: factorial and mixture (see Cornell [5]). While these two types of experiments may be fundamentally different, most of the same statistical tools applied to one experimental form may be applied to the other. The present study uses a combination of both types of experiments to create a single and more comprehensive experimental model. Such combinations of factorial/process components and mixture components are typically termed “crossed experiments.” In this case, the mixture components comprising the gaseous fuel blend are crossed with the process components of radial fuel split and percent pilot fuel injection at a fixed equivalence ratio and air preheat temperature.

Response Models. The final polynomial expressions for the responses (i.e., NO_x and CO) of the system are called the response models of the system. In the response models, each mixture component must always be represented since the total amount of a given mixture is always fixed. All other terms added to the response model—except those necessary to maintain hierarchy—were determined to have statistically significant (greater than 95%) impact upon the response of the system. Any terms added on a hierarchy basis was necessary to keep the model from becoming scale dependent.

The components and conditions used in the crossed experiment, and the terms that comprise response models are:

- Fixed Conditions
 - $\phi=0.52$
 - $T_{AIR}=730^{\circ}\text{F}$
- Mixture Components
 - A: CH₄=80%–100% (by volume)
 - B: C₂H₆=0%–15% (by volume)
 - C: C₃H₈=0%–20% (by volume)
- Process Components
 - D: % pilot=0%–5%
 - E: % centerbody=100%–40%

Using this combination of mixture and process components, a randomized test matrix consisting of only 72 experiments was generated. A sample of the test matrix is provided in Table 3. The total number of required measurements for this statistical approach is far less than would otherwise be required in a more traditional one-factor-at-a-time approach.

CO Response Model. The response model for the emissions of CO, in terms of actual factor units, is provided in Eq. (1). Note that the emissions of CO are measured in parts-per-million (ppm), and are corrected to 15% O₂ on a dry volume basis.

$$[\text{CO}]_{15\%} = 0.1029\text{A} + 0.1216\text{B} + 0.1511\text{C} + 1.215 \times 10^{-4}\text{AE} + 5.875 \times 10^{-4}\text{BE} + 8.258 \times 10^{-4}\text{CE}, \text{ (ppm)} \quad (1)$$

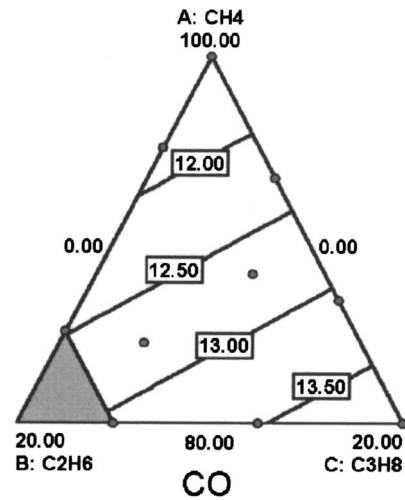


Fig. 6 Mixture space for CO at 100% centerbody (ppmv @ 15% O₂)

These results may also be depicted in graphical form as illustrated in Fig. 6. Note that the results shown in Fig. 6 are typical of other radial fuel splits, though the magnitude of the contour lines does change. Figure 6 indicates that the highest emissions of CO occur with the addition of propane with little contribution from ethane or methane.

A comparison of the predicted CO emissions from Eq. (1) to the measured emissions is provided in Fig. 7. This figure depicts the degree to which the Response Model can accurately predict the emissions of CO for a given set of conditions. The correlation coefficient (R-squared) of the model prediction compared to the measurements is 0.77. Note that the accuracy of the response model is independent of the reported accuracy of the individual measurements. The ability of the model to fit the measured data is dependent upon the judicious incorporation of components in the equation that are found to play the most significant role in impacting the response of interest, in this case CO.

In order to ascertain the relative importance/impact of each term in affecting the final response model for CO, the response model is presented coded form in Eq. (2).

$$[\text{CO}]_{15\%, \text{CODED}} = 11.14\text{A} + 12.17\text{B} + 13.09\text{C} + 0.36\text{AE} + 0.64\text{BE} + 0.79\text{CE} \text{ (ppm)} \quad (2)$$

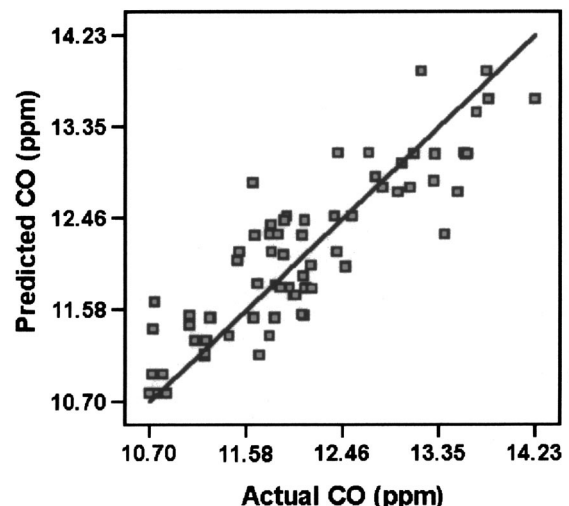


Fig. 7 Predicted versus actual values of CO (ppmv @ 15% O₂)

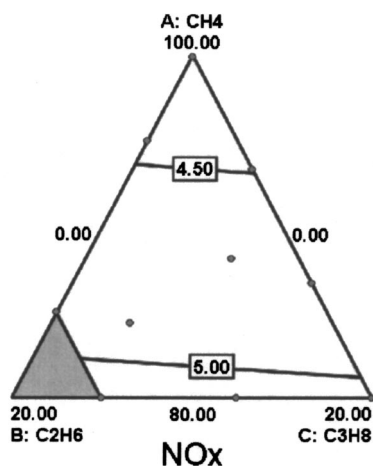


Fig. 8 Mixture space for NO_x at 100% centerbody (ppmv @ 15% O₂)

The coded form of the response model, unlike the response model in terms of actual components, is based on assignment of the range of each component to a -1 (for minimum component value) to $+1$ (for maximum component value). This allows the relative impact of changes in the components on the emissions of CO to be ascertained by simply comparing the coefficients of the various terms. For example, the term CE has a coefficient that is about twice that of AE, implying that the combined impact on the CO from a change in the percent centerbody and percent propane is twice the impact seen for a similar change made to the percent natural gas and percent centerbody. Additionally, all of the terms for the interactions in the coded form of the response model are positive, which implies that an increase in any of these components will lead to an increase in the CO.

Further examination of the coded terms for Eq. (2) leads to some interesting results. First, note that the coefficients for the three mixture components (A, B, & C) are substantially larger than any of the remaining terms. This implies that the emissions of CO—for the conditions under which this crossed experiment were conducted—are heavily influenced by the composition of the fuel. Increases in the radial fuel split (percent centerbody), for any fuel composition can also lead to an increase in the CO (given by the positive coefficients of the interacting terms), but the magnitude of increase is small relative to the overall magnitude of CO emissions. Additionally, there is no term showing an effect of the pilot fuel (D) in the coded response model. This means that the addition of pilot fuel, up to 5% of the total fuel volume, had no statistically significant effect on the CO for the conditions considered.

NO_x Response Model. The response model obtained for the emissions of NO_x in terms of actual components and ppm, corrected to 15% O₂ on a dry volume basis, is provided below in Eq. (3).

$$\begin{aligned}
 [\text{NO}_x]_{15\%} = & 0.0677\text{A} + 0.2548\text{B} - 0.1506\text{C} - 2.728 \times 10^{-3}\text{AD} \\
 & - 1.503 \times 10^{-3}\text{AE} - 3.765 \times 10^{-3}\text{BD} - 6.063 \\
 & \times 10^{-3}\text{BE} - 0.0193\text{CD} + 7.333 \times 10^{-3}\text{CE} + 1.248 \\
 & \times 10^{-5}\text{AE}^2 + 4.383 \times 10^{-5}\text{BE}^2 - 4.991 \times 10^{-5}\text{CE}^2 \\
 & + 7.138 \times 10^{-5}\text{ADE} + 9.078 \times 10^{-5}\text{BDE} + 2.401 \\
 & \times 10^{-4}\text{CDE}, \text{ (ppm)}
 \end{aligned} \quad (3)$$

As with CO, these results may be depicted in graphical form, such as the examples shown in Fig. 8 for 100% centerbody (CB),

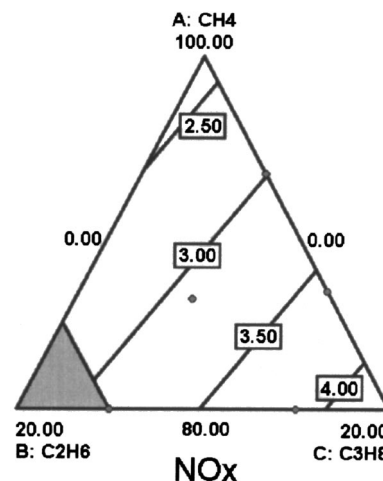


Fig. 9 Mixture space for NO_x at 70% centerbody (ppmv @ 15% O₂)

and Fig. 9 for 70% CB, both without pilot fuel. It should also be noted that most of the results obtained without pilot fuel are similar to Fig. 9 as the radial fuel split is varied from 100% CB. These figures indicate that the highest emissions of NO_x are obtained with the addition of propane, and that the role of the fuel composition on NO_x emissions also depends upon the radial distribution of the fuel. Both of these results are consistent with the findings of Flores et al. [2]. Figure 8 also indicates that the NO_x depends little on the fuel composition (i.e., the range in the contours is within experimental uncertainty). However, Fig. 9 indicates that the fuel composition does play a role with 70% CB and that, specifically, propane is the primary driver. This interaction between the radial fuel distribution and the fuel composition is captured in the model shown in Eq. (3).

A comparison of the predicted NO_x emissions from the response model to the actual NO_x emissions is provided in Fig. 10. The data depicted in Fig. 10 have a calculated R-squared value of 0.86.

As with the CO, additional insight may be obtained from an examination of the coded form of the response model for NO_x. The NO_x response model in coded form is as follows:

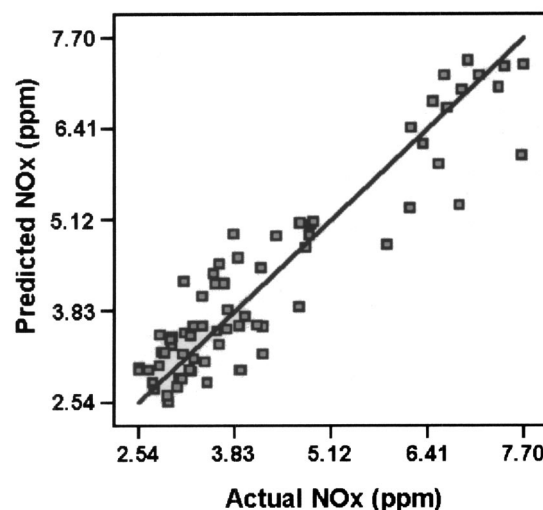


Fig. 10 Predicted versus actual values of NO_x (ppmv @ 15% O₂)

$$\begin{aligned}
[\text{NO}_x]_{15\%, \text{CODED}} = & 2.93\text{A} + 3.37\text{B} + 4.58\text{C} + 0.57\text{AD} + 1.27\text{AE} \\
& + 0.58\text{BD} + 1.19\text{BE} + 0.33\text{CD} + 1.58\text{CE} \\
& + 1.12\text{AE}^2 + 1.69\text{BE}^2 + 0.54\text{ADE} + 0.56\text{BDE} \\
& + 0.79\text{CDE} \quad (\text{ppm}) \quad (4)
\end{aligned}$$

From Eq. (4), the factors that play the most significant role in determining the level of NO_x emissions are the three gases used in the blend (i.e., the fuel composition). This is not to say that the mixture of the three gases is entirely dominant, but they have the most impact relative to the other terms. Also, propane again has the largest impact in affecting the level of NO_x emitted from the system, and not surprisingly, propane is followed by ethane and natural gas, respectively. This result is reflected by the coefficients of 4.58, 3.37, and 2.93 for propane, ethane, and natural gas, respectively.

Other interesting points may be made about the other terms that appear in the coded model, as well as a few points about terms that do not appear in the model. For instance, when comparing the coefficients for the terms AD and AE (the interaction between the methane ("A") and the components of % pilot ("D") or % centerbody ("E"), the percent centerbody plays a more important role than the percent pilot fuel injection, as characterized by AE's coefficient (1.27) versus AD's coefficient (0.57). This pattern repeats itself for ethane and propane. Just as telling is the fact that the percentage of the centerbody has squared terms with significant coefficient values for all three fuel components, whereas the percentage of the pilot has none. This result indicates a given composition of fuel will experience a significant rise in the level of NO_x emissions with an increase in the radial fuel split.

Lastly, the coded form of the NO_x model suggests that an interaction exists between the gaseous fuel composition and the addition of percent pilot and percent centerbody. This result makes sense in that increasing both the percentage of the pilot and centerbody leads to conditions where the largest amount of fuel is placed towards the center of the combustor premixer annulus (recall Fig. 4 and Fig. 5). This in turn leads to the highest "local" equivalence ratios achievable where thermal NO_x may be generated at increased levels.

Emissions Maps. To examine the ability of the applied statistical methodology to accurately predict the trends of the present system for different fuel compositions, several emissions maps were generated. Maps are provided for the extremes in fuel compositions shown in Table 2, which were also used in Flores et al. [2] for a different type of injector. The emissions maps, with no pilot fuel, are shown in Fig. 11.

The emissions of NO_x and CO described on the maps in Fig. 11 are corrected to 15% O_2 , and are plotted as a function of equivalence ratio and radial fuel split (described in terms of the percentage of fuel injected radially outwards from the centerbody). Also, the white regions on the lean side of each plot reflect regions beyond the LBO limits, where stable combustion could not be achieved. Note that the fuel split between the centerbody and wall injectors is of the remaining fuel *after* any pilot fuel is subtracted.

CO Emissions Maps. Recall that the statistical experimentation performed was for a fixed equivalence ratio of 0.52. However, several of the key observations made with respect to the CO appear to be valid for the CO emissions maps provided in Fig. 11. One finding of the crossed experiments performed was that the emissions of CO were almost exclusively determined by the composition of the fuel, and that the radial fuel split had little impact on CO. A slight increase in the CO with an increase in the radial fuel split is noted, but that shift is small compared to the overall magnitude of the CO emissions as determined by the equivalence ratio and fuel composition. This is also illustrated in the CO plots of Fig. 11 by the near vertical contour lines that only appear to shift to slightly higher levels of CO as the radial fuel split is increased and approaches a value of 100% centerbody.

Likewise, the affect of the Pilot fuel on the emissions of CO is almost negligible, failing to rise above the experimental error (± 0.53 ppm) for these results. An emissions map for CO showing the impact of the pilot fuel is, therefore, not provided, given that there is little difference between such maps and their respective counterparts in Fig. 11. This apparent lack of CO sensitivity to the overall fuel split is primarily due to the residence time of the fuel and air mixture in the combustion liner, and not the level of mixing generally achieved, since a similar result was obtained in Flores et al. [2] with significantly poorer fuel and air mixing. Much of the CO that is produced, therefore, is given sufficient time to oxidize and form CO_2 , and the overall fuel splits used in the crossed experiments (40–100% centerbody and 0–5% pilot fuel) does little to change this.

One inconsistency is observed between the maps and the response model. The CO response model indicates that the lowest emissions were for the fuel comprised of 100% natural gas, and generally highest with the addition of propane. The CO results of Fig. 11 indicate, however, that it is actually the fuel composition with the significant levels of ethane that yield the lowest levels of CO. However, this difference may be attributed to the experimental error of the system as well as the scatter in the data comprising the CO response model (recall that Fig. 7 had an R-squared value of 0.77).

NO_x Emissions Maps. One of the most interesting observations made via the NO_x response model was that the impact of the radial fuel split is second order in nature. Whether or not this observation holds up against the data represented in Fig. 11 is not immediately obvious. As such, some of the NO_x data points from Fig. 11(a) are graphed in Fig. 12. Indeed, Fig. 12 indicates that the emissions of NO_x increase rapidly with the radial fuel split as it approaches 100% centerbody. Similar results to those provided in Fig. 12 are also obtained for the other two fuel compositions mapped in Fig. 11.

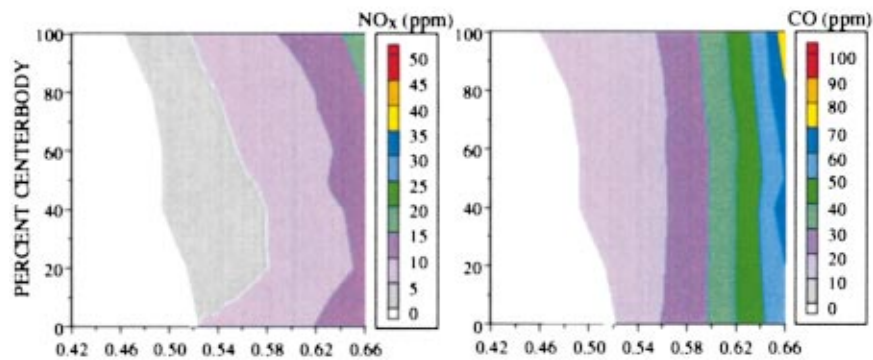
It was previously mentioned that the NO_x response model indicated that the addition of propane would lead to the largest increase in the emissions of NO_x over the baseline fuel, followed by an increase in NO_x resulting from the addition of ethane. This observation, however, is not directly supported with the NO_x results provided in Fig. 11, the addition of ethane led to lower emissions of NO_x than measured with the baseline fuel of 100% natural gas. A similar result was noted in the results of Flores et al. [2] where some of the same fuel blends were examined on a different system at an air preheat temperature of 700 K.

The apparent discrepancy between the predicted behavior from Eq. (3) and the detailed maps shown in Fig. 11 is attributed to (1) the use of a single equivalence ratio for the development of the model, (2) the inability of the model to capture the subtle fuel effects revealed in the more detailed tests (i.e., R^2 of 0.86), and (3) the presence of experimental error associated with the measurements. Careful examination of Fig. 9 does indicate that the response model predicts marginal increases in NO_x over the baseline fuel with significant amounts of ethane for at least some of the fuel split conditions.

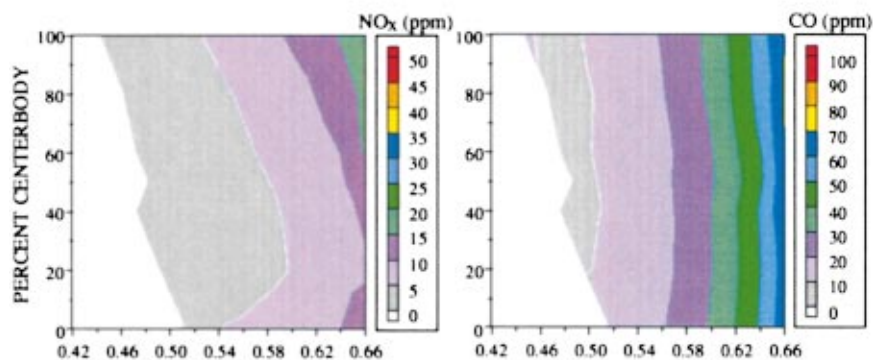
Another parameter that is predicted to have an effect on the emissions of NO_x was the inclusion of fuel through the pilot circuit. The NO_x response model indicated that the addition of pilot fuel would play a role in the formation of NO_x . For the cases with 20% propane, there is a significant increase in the formation of NO_x at radial fuel splits greater than 50% centerbody (when compared for the case without pilot fuel injection). For the other two fuel blends, the impact of the pilot fuel is amplified as the radial fuel split increases above 60–70% centerbody (for the range of equivalence ratios plotted), supporting the ADE, BDE, and CDE terms listed in the coded form of the NO_x response model.

This result is due to the fact that increasing the radial fuel split directs additional fuel to the centerline of the combustor premixer

a) 100% Natural Gas



b) 85% Natural Gas & 15% Ethane



c) 80% Natural Gas & 20% Propane

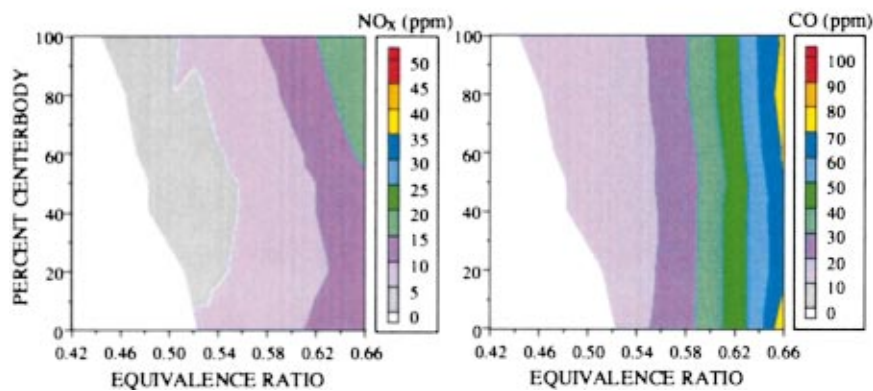


Fig. 11 NO_x and CO emissions maps for three fuel compositions without pilot injection (ppmv @ 15% O_2)

annulus, where it begins to approximate a relatively well-mixed axial injector (recall Fig. 4 and Fig. 5). Add to this the effect of another 5% of the total volumetric flow rate in a manner that is not as well mixed as the radially injected fuel (but certainly not a diffusion type injector due to the mixing distance). Altogether, this allows the pilot fuel stream to magnify/compound its apparent effect on the emissions of NO_x for the system.

Potential NO_x Formation Pathways. Given the significance and emphasis generally placed on the emissions of NO_x relative to the emissions of CO, a brief discussion as to the likely sources of the NO_x emissions is warranted. Furthermore, given the relatively low equivalence ratios (and relatively low reaction temperatures that result) in Fig. 11 and Fig. 12, it is unlikely that the thermal NO_x pathway is responsible for all of the measured NO_x ,

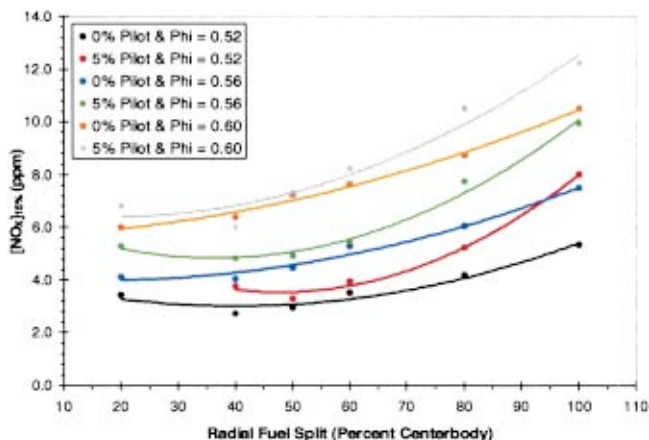


Fig. 12 NO_x versus radial fuel split for 100% natural gas (ppmv @ 15% O_2)

leading to a potentially significant contribution from other NO_x formation pathways. Moreover, the question arises about the specific role that fuel composition may have in the various nonthermal NO_x formation mechanisms.

In order to better understand the role of the fuel composition on the NO_x formed, and to gain insight into the potential for nonthermal NO_x pathways to contribute, some limited chemical kinetic calculations were conducted with a sensitivity analysis. Applying Chemkin's Senkin software package, [6], and GRI Mech 3.0 to several selected conditions appears to suggest that it is likely that the NO_x generated was derived via the nitrous oxide (N_2O) pathway in conjunction with the thermal NO_x pathway, with the N_2O pathway likely responsible for a substantial portion (greater than 75%) of the NO_x measured. This is not surprising, however, when one considers that while thermal NO_x may be present, it requires high reaction temperatures, typically above 1900 K, to become dominant. Below this temperature, other NO_x pathways may play a significant role. Furthermore, the percentage and amount of NO_x formed from the N_2O pathway appears to increase with the addition of propane.

Another pathway, the prompt NO_x pathway, likely contributes little to the overall emissions of NO_x from the system. Prompt NO_x is unlikely to produce as much NO_x as the thermal and N_2O pathways. Prompt NO_x can continue to play a role at reaction temperatures below 1800 K; however, in lean premixed systems, it has been indicated that prompt NO_x is not a major source of NO_x until the equivalence ratio exceeds values greater than approximately 0.65 (Steele et al. [7]). The N_2O Pathway, on the other hand, has been found in other studies to play a significant role in the overall levels of NO_x emitted from lean premixed reactions with equivalence ratios less than 0.80, [8].

As such, a significant portion of NO_x measured and shown above must be a result of the N_2O pathway. This is most likely for the cases with relatively well mixed fuel and air that create few, if any, "hot spots" that may accelerate the thermal NO_x mechanism. However, the proportional contribution of the N_2O pathway is reduced when compared to thermal NO_x pathway at conditions that are not well mixed, or have areas of locally high equivalence ratio (like those attained with high radial fuel splits and pilot fuel addition), where the production of thermal NO_x may begin to significantly increase (see Fig. 12). Note that Fig. 4 and Fig. 5 clearly indicate that while the overall equivalence ratios from Fig. 11 may be low, the significant amount of fuel placed at the centerline of the combustor premixer annulus leads to areas of relatively high local equivalence ratio and temperature where thermal NO_x becomes a stronger component of the overall NO_x emissions.

Summary and Conclusions

The results presented illustrate the potential for a design of experiments (DoE) testing methodology to provide a powerful tool to researchers conducting a variety of studies; in this case, a study focused in fuel compositional effects. The two general types of DoE methodologies, factorial and mixture, were combined in a type of "hybrid" methodology called a "crossed" methodology. This hybrid methodology sought to combine process and mixture components to examine the impact upon the selected responses of the system, which were the emissions of NO_x and CO. The effectiveness of the crossed methodology was compared to additional data obtained and presented in the form of emissions maps for both NO_x and CO.

The results indicate that

- the crossed model approach captures the majority of the emissions behavior noted in the emissions maps, and has generated an empirical expression that relates the emissions to both fuel composition and fuel distribution. The only significant discrepancy of this approach was its failure to capture the subtle reduction in emissions obtained with the addition of significant amounts of ethane.
- CO formation is dominated by the fuel composition with higher hydrocarbons leading to higher CO levels. The fuel distribution and operation of the pilot has little influence on the CO levels in this system.
- NO_x formation in the present system is dependent upon the fuel composition, but also exhibits subtle dependencies upon the fuel distribution and the use of the pilot. The crossed model generates an expression that captures the coupling between the fuel distribution and the composition.
- the composition of the fuel had a negligible impact upon the curvature of the radial fuel distribution profile entering the combustor.
- the emissions of CO were noted to be relatively insensitive to the radial fuel split for both the present system as well as the system previously used in Flores et al. [2] which achieved relatively poor fuel and air mixing.
- limited kinetic calculations with a sensitivity analysis appears to suggest that the nitrous oxide pathway predominantly responsible for the measured emissions of NO_x .

Acknowledgments

The authors acknowledge the support of the U.S. Department of Energy Advance Turbine Systems program and the South Carolina Institute for Energy Studies (SCIES). The contributions of Mr. Junhua Chen, Dr. Jack Brouwer, and Mr. Steve Hill are gratefully appreciated.

References

- [1] Goy, C. J., Moran, A. J., and Thomas, G. O., 2001, "Autoignition Characteristics of Gaseous Fuels and Representative Gas Turbine Conditions," ASME Paper No. 2001-GT-0051.
- [2] Flores, R. M., Miyasato, M. M., McDonell, V. G., and Samuelsen, G. S., 2001, "Response of a Model Gas Turbine Combustor to Variation in Gaseous Fuel Composition," ASME J. Eng. Gas Turbines Power, **123**(4), pp. 824–831.
- [3] Liss, W. E., Thrasher, W. H., Steinmetz, G. F., Chowdiah, P., and Attari, A., 1992, "Variability of Natural Gas Composition in Select Major Metropolitan Areas of the United States," GRI-92/0123, Mar.
- [4] Box, G. E. P., Hunter, W. G., and Hunter, J. S., 1978, *Statistics for Experiments*, John Wiley and Sons, New York.
- [5] Cornell, J. A., 1990, *Experiments With Mixtures: Designs, Models, and the Analysis of Mixture Data*, 2nd Ed., John Wiley and Sons, New York.
- [6] Kee, R. J. et al., 1999, *Chemkin Collection*, Release 3.5, Reaction Design, Inc., San Diego, CA.
- [7] Steele, R. C., Tonouchi, J. H., Nicol, D. G., Horning, D. C., Malte, P. C., and Pratt, D. T., 1998, "Characterization of NO_x , N_2O , and CO for Lean-Premixed Combustion in a High-Pressure Jet-Stirred Reactor," ASME J. Eng. Gas Turbines Power, **120**, pp. 303–310.
- [8] Turns, S. R., 1996, *An Introduction to Combustion: Concepts and Applications*, McGraw-Hill, New York.
- [9] Meier, J. G., Hung, W. S. Y., and Sood, V. M., 1999, "Development and Application of Industrial Gas Turbines for Medium-BTU Gaseous Fuels," ASME J. Eng. Gas Turbines Power, **108**, pp. 182–190.

C. M. Spadaccini

Gas Turbine Laboratory,
Department of Aeronautics and Astronautics,
Massachusetts Institute of Technology,
Cambridge, MA 02139

A. Mehra

D-STAR Engineering Corporation,
Shelton, CT 06484

J. Lee

Gas Turbine Laboratory,
Department of Aeronautics and Astronautics,
Massachusetts Institute of Technology,
Cambridge, MA 02139

X. Zhang

Department of Manufacturing Engineering
and
Franhofer USA Center for Manufacturing
Innovation,
Boston University,
Boston, MA 02115

S. Lukachko

I. A. Waitz

Gas Turbine Laboratory, Department of
Aeronautics and Astronautics,
Massachusetts Institute of Technology,
Cambridge, MA 02139

High Power Density Silicon Combustion Systems for Micro Gas Turbine Engines

As part of an effort to develop a microscale gas turbine engine for power generation and micropropulsion applications, this paper presents the design, fabrication, experimental testing, and modeling of the combustion system. Two radial inflow combustor designs were examined; a single-zone arrangement and a primary and dilution-zone configuration. Both combustors were micromachined from silicon using deep reactive ion etching (DRIE) and aligned fusion wafer bonding. Hydrogen-air and hydrocarbon-air combustion were stabilized in both devices, each with chamber volumes of 191 mm³. Exit gas temperatures as high as 1800 K and power densities in excess of 1100 MW/m³ were achieved. For the same equivalence ratio and overall efficiency, the dual-zone combustor reached power densities nearly double that of the single-zone design. Because diagnostics in microscale devices are often highly intrusive, numerical simulations were used to gain insight into the fluid and combustion physics. Unlike large-scale combustors, the performance of the microcombustors was found to be more severely limited by heat transfer and chemical kinetics constraints. Important design trades are identified and recommendations for microcombustor design are presented. [DOI: 10.1115/1.1586312]

Introduction

Recent advances in silicon microfabrication techniques and microelectromechanical systems (MEMS) have led to the possibility of a new generation of micro heat engines for power generation and micro air-vehicle propulsion applications. Epstein et al. [1] and Groshenry [2] have reported the design for a silicon-based, micro gas turbine generator capable of producing 10–50 Watts of power in a volume less than 1 cm³ while consuming seven grams of fuel per hour. This represents a tenfold increase in power density over the best available batteries. Like their larger counterparts, an engine of the type shown in Fig. 1 requires a high-temperature combustion system to convert chemical energy into fluid thermal and kinetic energy.

This paper seeks to elucidate the underlying physics unique to microscale combustion systems through an experimental, computational, and analytical investigation. These studies have resulted in the identification of critical design trades and recommendations for microcombustor design.

The work presented here is part of a broader effort to fabricate and operate this microscale gas turbine engine. The development strategy involves concurrently developing component technologies such as microcompressors, gas bearings, and microcombustors.

All of these individual components have undergone independent testing and the first build of the engine has recently been completed.

The objective of a separate combustor test rig is to study combustor performance without the complexity of the full engine. The designs presented here retain most of the important functionality and boundary conditions of the engine. This includes overall flow path and geometry, inlet swirl angle, combustor pressure, exit area, and back pressure. The primary difference between the combustor rigs and the actual engine is the lack of a rotor. The tests are intended to explore the operating space of these combustors in an effort to determine if they provide adequate residence time, turbine inlet temperature, mass flow, and pressure ratio for the engine cycle. The final engine design will reflect these findings as well as those from other component experiments and analyses.

Section 2 discusses the challenges of combustion in a microscale environment and overviews previous work towards the development of silicon-based systems. The design and testing methodology for two microcombustion devices are introduced in Section 3. In Section 4, results from the experiments are presented, complemented by results from computations and analysis. The results are synthesized in Section 5 as an operating space and in Section 6, the resulting design trades are reviewed. Section 7 summarizes and concludes the paper.

2 Microcombustion Challenges

The functional requirements of a microcombustor are similar to those of a conventional gas turbine combustor. These include the

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Amsterdam, The Netherlands, June 3–6, 2002; Paper No. 2002-GT-30082. Manuscript received by IGTI, December 2001, final revision, March 2002. Associate Editor: E. Benvenuti.

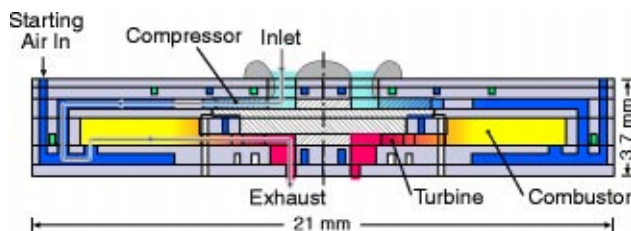


Fig. 1 Baseline engine schematic

efficient conversion of chemical energy to fluid thermal and kinetic energy with low total pressure loss, reliable ignition, and wide flammability limits. However, the obstacles to satisfying these requirements are different for a microscale device. As first described by Waitz et al. [3] a microscale combustor is more highly constrained by inadequate residence time for complete combustion and high rates of heat transfer from the combustor. Microcombustor development also faces unique challenges due to material and thermodynamic cycle constraints. These constraints are reviewed in the following sections.

2.1 Time-Scale Constraints. For the energy conversion applications we are interested in, power density is the most important metric. As shown in Table 1, the high-power density of a microcombustor directly results from high mass flow per unit volume. Since chemical reaction times do not scale with mass flow rate or combustor volume, the realization of this high-power density is contingent upon completing the combustion process within a shorter combustor through-flow time.

This fundamental time constraint can be quantified in terms of a homogeneous Damköhler number; the ratio of the residence time to the characteristic chemical reaction time.

$$Da_h = \frac{\tau_{\text{residence}}}{\tau_{\text{reaction}}} \quad (1)$$

To ensure a Da_h greater than unity (and complete combustion), a designer of a microcombustor can either increase the flow residence time or decrease the chemical reaction time. The characteristic combustor residence time is given by the bulk flow through the combustor volume.

$$\tau_{\text{residence}} \approx \frac{\text{volume}}{\text{volumetric flow rate}} = \frac{VP}{\dot{m}RT} \quad (2)$$

Residence time can be increased by increasing the size of the chamber, reducing the mass flow rate, or increasing the operating pressure. A chemical reaction time can be approximated by an Arrhenius type expression.

Table 1 A comparison of the operating parameters and requirements for a microengine combustor with those estimated for a conventional GE90 combustor

	Conventional Combustor	Microcombustor
Length	0.2 m	0.001 m
Volume	0.073 m ³	6.6 × 10 ⁻⁸ m ³
Cross-sectional area	0.36 m ²	6 × 10 ⁻⁵ m ²
Inlet total pressure	37.5 atm	4 atm
Inlet total temperature	870 K	500 K
Mass flow rate	140 kg/s	1.8 × 10 ⁻⁴ kg/s
Residence time	~7 ms	~0.5 ms
Efficiency	>99%	>90%
Pressure ratio	>0.95	>0.95
Exit temperature	1800 K	1600 K
Power density	1960 MW/m ³	3000 MW/m ³

Note: residence times are calculated using inlet pressure and an average flow temperature of 1000 K

$$\tau_{\text{reaction}} \approx \frac{[\text{fuel}]_0}{A[\text{fuel}]^a[\text{O}_2]^b e^{-E_a/RT_0}} \quad (3)$$

Reaction time is primarily a function of fuel properties and the mixture temperature and pressure.

Since high-power density requirements mandate high mass flow rates through small chamber volumes, the mass flow rate per unit volume can not be reduced without compromising device power density. Hence, there is a basic tradeoff between power density and flow residence time.

$$\text{Power density} \propto \frac{\dot{m}}{V} \propto \frac{\dot{m}_f \text{LHV}}{V} \propto \frac{\rho}{\tau_{\text{residence}}} \quad (4)$$

For a given operating pressure (and thus density), and assuming a Da_h of unity, reducing the chemical reaction time and thus required residence time is the only means of ensuring complete combustion without compromising the high power density of the device.

Mixing time scales are also critical in microcombustion systems. Due to the small length scales there is little time for fuel-air mixing and inadequate mixing can also lead to chemical inefficiency.

2.2 Heat Transfer Effects and Fluid-Structure Coupling

Energy loss due to heat transfer at the walls of the combustion chamber in a conventional gas turbine is typically neglected. However, for a microcombustor this is an important factor. The surface-area-to-volume ratio for a microscale combustor is approximately 500 m⁻¹, or two orders of magnitude larger than that of a typical combustor.

Waitz et al. [3] have shown that the ratio of heat lost to that generated scales with the hydraulic diameter as follows:

$$\frac{E''}{\dot{E}} \propto \frac{1}{d_h^{1.2}} \quad (5)$$

The hydraulic diameter of a microcombustor is on the order of millimeters, hundreds of times smaller than that of a typical combustor. Therefore, the ratio of heat lost to that generated may be as much as two orders of magnitude greater than that of a large-scale combustor.

The effects of this large surface heat loss on combustion are twofold. First, large thermal losses have a direct impact on overall combustor efficiency. Therefore, typical large-scale combustor efficiencies of greater than 99% are not feasible. Second, they can increase kinetic times and narrow flammability limits through lowering reaction temperatures. This can exacerbate the constraints of short residence time.

2.3 Recirculation Zones. Another important factor to consider in microcombustor development is the design of recirculation zones for ignition and flame stability. To maintain high power density a flame-holding structure in a microcombustion system must not only stabilize the flame, but must do it in a manner that efficiently uses the space within the combustor. Large recirculation zones or designs that do not initiate reactions rapidly in all portions of the flow reduce the power density that can be achieved in the device.

2.4 Material Constraints. There are also several material constraints imposed upon a silicon microcombustion system. The most critical requirement is a wall temperature limit of less than 950 K. At temperatures above this level, silicon begins to soften and lose its structural integrity. However, high surface heat transfer and the high thermal conductivity of silicon are beneficial in this case. Combustor wall temperatures can be kept below the 950 K requirement by conduction of heat through the structure to the ambient. In addition, the rotating components of the micro-engine must maintain even lower wall temperatures, below 900 K, due to creep considerations.

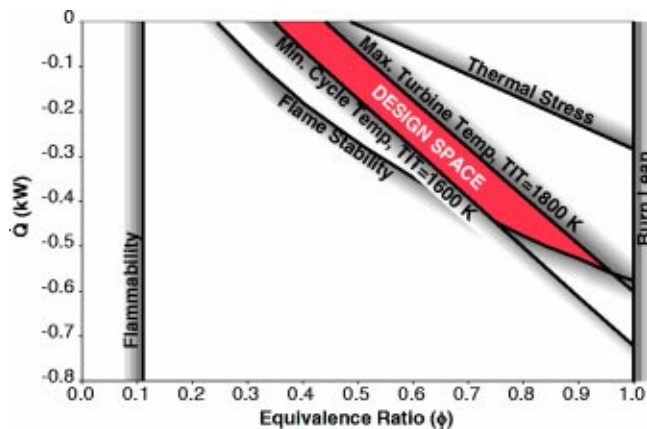


Fig. 2 Microcombustor design space

2.5 Design Space. While time scale, heat transfer, and material constraints are important, the most limiting are those imposed by the engine thermodynamic cycle. Figure 2 shows a typical design space for a hydrogen-fueled microscale combustor. Equivalence ratio is plotted on the x-axis versus heat loss on the y-axis. A turbine inlet temperature between 1600 K and 1800 K is required for the cycle proposed for this engine. In Fig. 2 the flame stability limit represents the stable operation of a nonadiabatic perfectly stirred reactor (PSR). The thermal stress constraint indicates the material limits of the all silicon structure of a micro-machined combustion system. The design space is further bounded by the desire to burn lean and the flammability limits of hydrogen-air combustion.

2.6 Previous Work. Mehra and Waitz [4] were the first to develop a silicon, microfabricated combustor compatible with a realistic engine geometry. This three-wafer combustor, shown in Fig. 3, was 0.066 cm^3 in volume and was designed to operate using lean, premixed, hydrogen-air combustion. The device also had the ability to operate with fuel injection.

The combustor was tested over a range of equivalence ratios spanning from 0.4 to 1.6 for a fixed mass flow rate of 0.045 g/s and atmospheric pressure. This resulted in a residence time of approximately 0.5 ms. For premixed hydrogen-air operation, exit gas temperatures in excess of 1800 K were achieved with combustor efficiencies of up to 70%. For nonpremixed operation, peak exit gas temperatures were approximately 100 K lower with efficiencies above 65%. This reduced performance indicated that there was inadequate fuel-air mixing upstream of the combustion chamber.

Low combustor efficiencies were attributed to poor thermal isolation of the device. Chemical conversion was estimated at nearly 100% while heat transfer from the combustion chamber to the surroundings lowered the overall efficiency. The power density of the device was approximately 1200 MW/m^3 . The device also maintained its structural integrity over tens of hours of operation. References [4] and [5] contain more detail regarding this combustor.

3 Experimental Approach

To improve upon past microcombustor performance and to further understand the physics of these microscale systems, two new devices were fabricated. Both are compatible with the micro-engine design shown in Fig. 1 and both combustors were con-

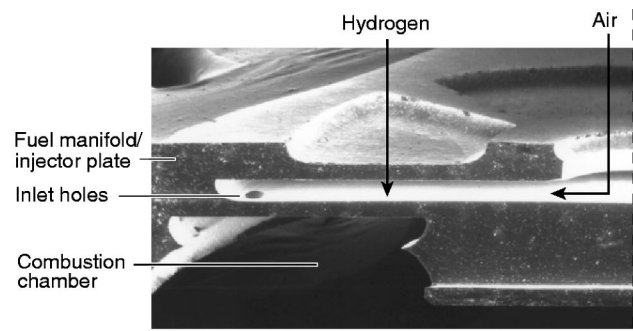
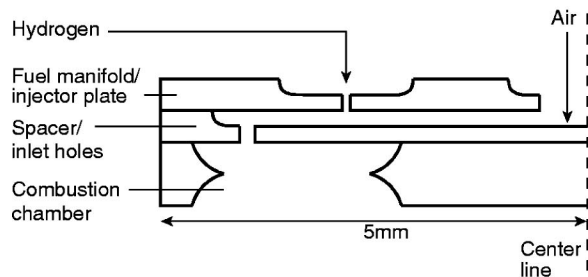


Fig. 3 Schematic and SEM of 0.066 cm^3 microcombustor

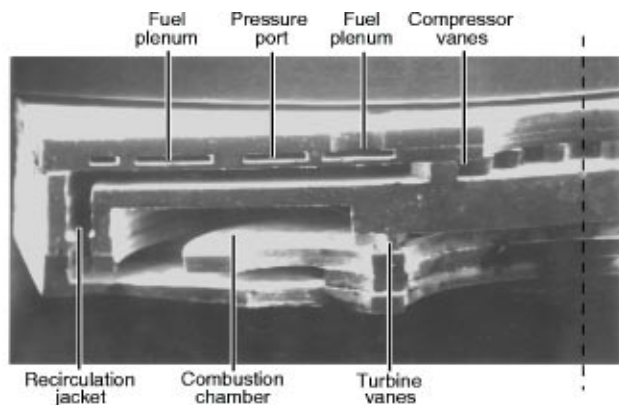
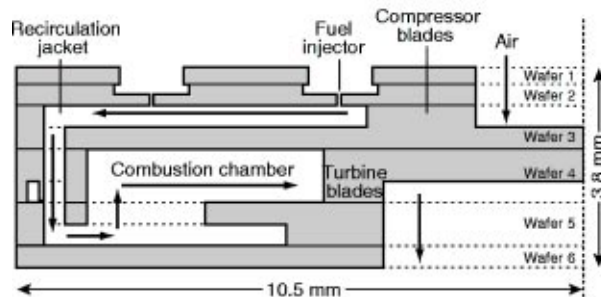


Fig. 4 Schematic and SEM of six-wafer combustor

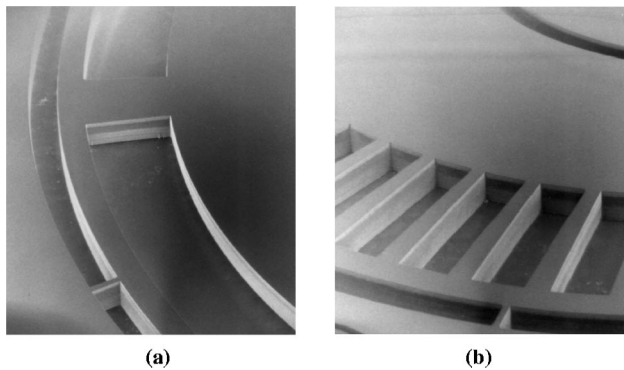


Fig. 5 SEMs of combustion chamber inlets

structed using the microfabrication techniques previously developed, as well as some new advances. The details of the design and fabrication can be found in Refs. [5] and [6].

3.1 Six-Wafer Microcombustor. A schematic of the six-wafer microcombustor is shown in Fig. 4 alongside a scanning electron micrograph (SEM) of a cross-section of the actual device. Air, or premixed fuel-air, enters the device axially followed by a 90 deg turn prior to entering the combustor. For this nonrotating test device, swirl vanes were included to replicate the compressor exit flow angle of 80 degs. The flow then passes through a duct which wraps around the combustion chamber. This passage is referred to as the *cooling jacket* and is intended to thermally isolate the combustion chamber and cool the inner walls, as well as act as a simple recuperator for preheating the reactants. This is also where fuel is injected in the non-premixed mode of operation. A set of three fuel injection ports were included to evaluate fuel-air mixing requirements.

The fuel-air mixture is then burned in the combustion chamber. Two types of chamber inlets were designed to create different flame holding recirculation zones. The first is a simple annular inlet 1.2 mm wide. The second is an array of 60 slots 2.2 mm long intended to create multiple small recirculation zones for more rapid and uniform ignition of the incoming flow. SEMs of both inlets can be seen in Fig. 5. Finally, the combustion products are passed through the turbine nozzle guide vanes and exit axially from the device.

3.2 Dual-Zone Microcombustor. The dual-zone microcombustor is a modification to the six-wafer geometry, but represents a significant change in operating mode. A schematic of this device is shown in Fig. 6. A series of holes through the inner wall have been created connecting the upper cooling jacket to the combustion chamber. These holes allow inlet air to bleed into the

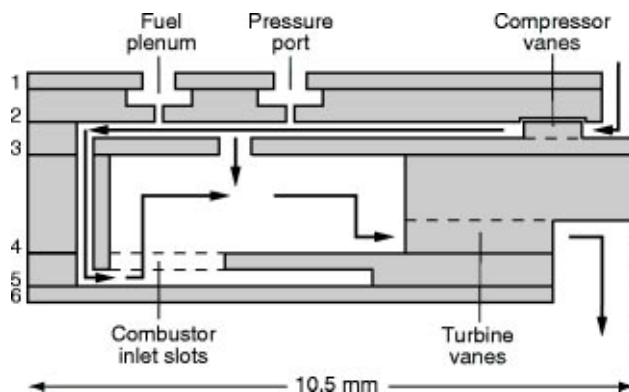


Fig. 6 Schematic of the dual-zone combustor

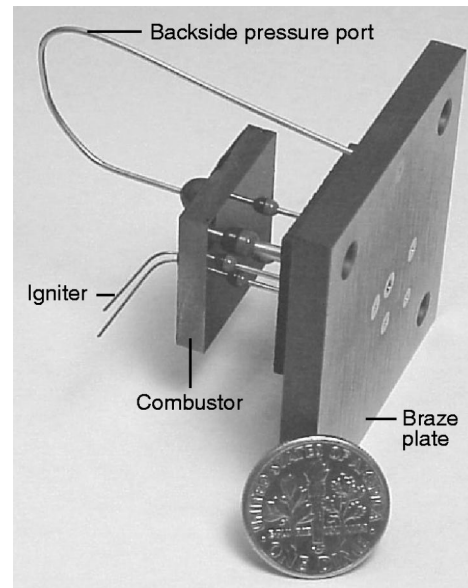


Fig. 7 Fully packaged microcombustor

combustion chamber. This dilution air serves two purposes; first, it splits the combustion chamber into two zones, and second, it dilutes the hot combustion products reducing their temperature to the desired turbine inlet temperature of 1600 K. This is similar to conventional large-scale combustors. In the primary zone, the fuel-air mixture is burned near stoichiometric conditions. By burning at higher temperature, kinetic rates increase and reaction time decreases effectively increasing Da_h . The dilution jet also serves to strengthen recirculation zones and improve stability. The combustion chamber inlet for the dual-zone combustor is slotted as shown in Fig. 5(b). Due to the nature of this combustor's operation, premixed combustion is not feasible and fuel injection occurs just downstream of the dilution holes in the cooling jacket.

The dilution holes were designed to split the total mass flow in half, establishing a stoichiometric primary zone at the design point conditions. The overall equivalence ratio for the combustor in this case is 0.5. However, at off-design conditions, the flow split deviates from the desired 50% dilution air and the primary zone will no longer operate at $\phi=1.0$. As mass flow rate increases, the percentage of total flow used for dilution decreases. For a fixed overall fuel/air ratio, this results in a lean primary zone.

Two variations of the dual-zone microcombustor were fabricated to test the effect of primary zone size. These variations differed in the radial locations of dilution holes. Combustors with dilution holes located at a 5.9 mm radius and a 7.0 mm radius were constructed.

3.3 Experimental Setup and Diagnostics. Prior to testing, the microcombustors were packaged into a suitable test rig. In order to connect the device's microscale fluid channels to a macroscale feed system, a glass bead interconnect scheme was developed by Mehra [5]. Small kovar tubing was hermetically sealed to the silicon with glass beads and brazed to a larger invar plate for connection to conventional fittings. A fully packaged device is shown in Fig. 7. A more detailed description of this process can be found in references [5–7].

Due to the microscale of the devices, it is difficult to obtain nonintrusive measurements. Therefore, diagnostics were limited. Exit gas temperature was measured using a 0.010 in. sheathed type K thermocouple. Because of the large temperature gradients along the length of the wire, an error analysis for the thermal conductivity, radiative emissivity, and calibration drifts predicted uncertainties up to ± 130 K. A wall temperature measurement was also obtained with the same type thermocouple and an uncertainty

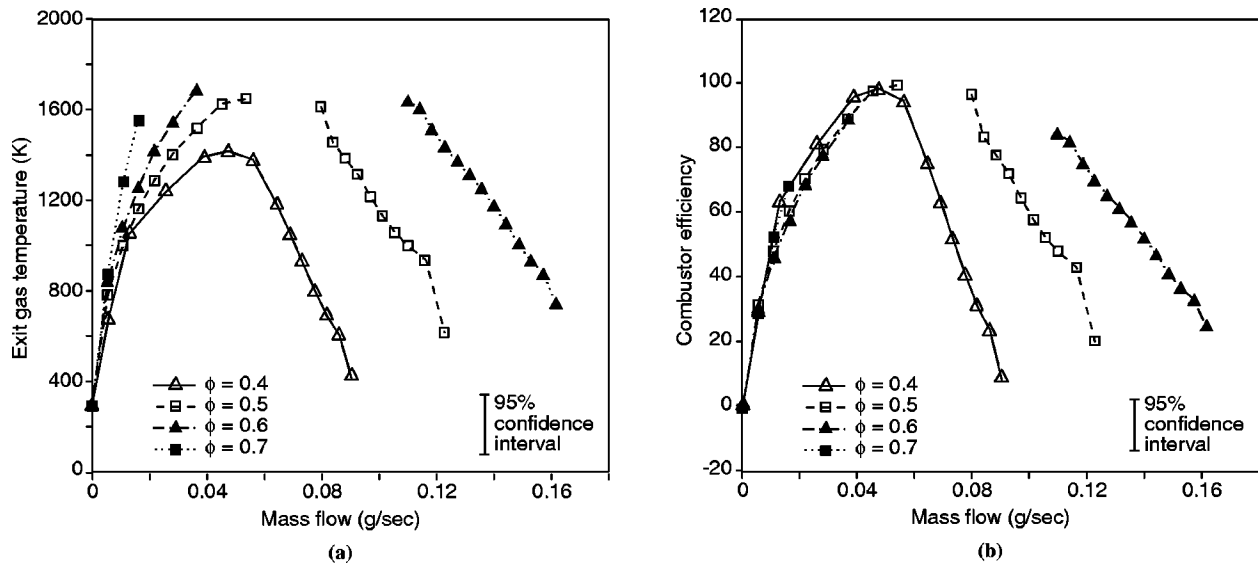


Fig. 8 Plots of exit gas temperature and efficiency for annular six-wafer combustor with premixed hydrogen-air

of ± 12 K. In addition to the temperature diagnostics, pressure was measured upstream in the cooling jacket and in the combustion chamber itself. Flow control was accomplished with several MKS Type 1359C flow controllers with $\pm 0.5\%$ full-scale accuracy (± 0.05 SLM for air and ± 0.005 SLM for fuel). A detailed uncertainty analysis can be found in Ref. [5].

4 Results and Analysis

Overall combustor efficiency is defined as

$$\eta_c = \frac{(\dot{m}_a + \dot{m}_f) h_2 - \dot{m}_a h_1}{\dot{m}_f h_f} \quad (6)$$

where station (1) refers to the combustor inlet and station (2) is the combustor exit. The combustor efficiency can be written as the product of two subefficiencies: a *chemical efficiency*, and a *thermal efficiency*. These two efficiencies can be written as

$$\eta_{\text{chem}} = \frac{[(\dot{m}_a + \dot{m}_f) h_2 - \dot{m}_a h_1] + \dot{Q}_{\text{loss}}}{\dot{m}_f h_f} = \frac{\text{total enthalpy released}}{\text{maximum enthalpy release possible}} \quad (7a)$$

$$\eta_{\text{therm}} = \frac{(\dot{m}_a + \dot{m}_f) h_2 - \dot{m}_a h_1}{[(\dot{m}_a + \dot{m}_f) h_2 - \dot{m}_a h_1] + \dot{Q}_{\text{loss}}} = \frac{\text{enthalpy rise of fluid}}{\text{total enthalpy released}} \quad (7b)$$

4.1 Baseline Six-Wafer Hydrogen Tests. Baseline six-wafer microcombustor tests were performed with premixed hydrogen-air in the annular inlet geometry. Figure 8 shows plots of combustor exit temperature and efficiency for constant equivalence ratio.

The combustor produced exit gas temperatures in excess of 1600 K for a mass flow rate of 0.11 g/s and an operating pressure of 1.13 atm. The overall efficiency in this case exceeded 90% and the power density was approximately 1100 MW/m³. The break in the $\phi=0.5$ and $\phi=0.6$ exit temperature and efficiency curves is due to a lack of measurement capabilities for gas temperatures over 1600 K. The $\phi=0.7$ curve terminates at a mass flow rate of 0.015 g/s due to upstream burning in the cooling jacket.

The inability of the device to achieve the design point operation of 0.36 g/s and 3 atm and the general shape of the exit temperature and efficiency curves can be explained by examining the change

in Da_h over a constant equivalence ratio operating line. Using the previously defined residence time and a simple one step mechanism for hydrogen-air combustion

$$\tau_{\text{reaction}} = \frac{[H_2]_0}{A[H_2]^2[O_2]e^{-E_a/RT}} \quad (8)$$

with $E_a = 10950$ cal/mole and $A = 1.62 \times 10^{18}$, an estimate of Da_h can be obtained, [5]. Figure 9 shows a plot of reaction time and residence time along with Da_h versus mass flow rate. Residence time decreases more rapidly than reaction time, resulting in a Da_h less than one. This leads to chemical inefficiency at high mass flow rates as well as flame blowout prior to choking the nozzle guide vanes and reaching the design point conditions.

Analysis by Mehra [5] using a one-dimensional heat transfer model to predict heat loss from the combustor indicates that thermal inefficiencies dominate low mass flow operation, while at high mass flows performance is limited by chemical inefficiency.

4.2 Comparison of Inlet Geometry. Combustor efficiency for premixed hydrogen-air tests in the slotted-inlet geometry is plotted in Fig. 10. Performance is similar to the annular inlet

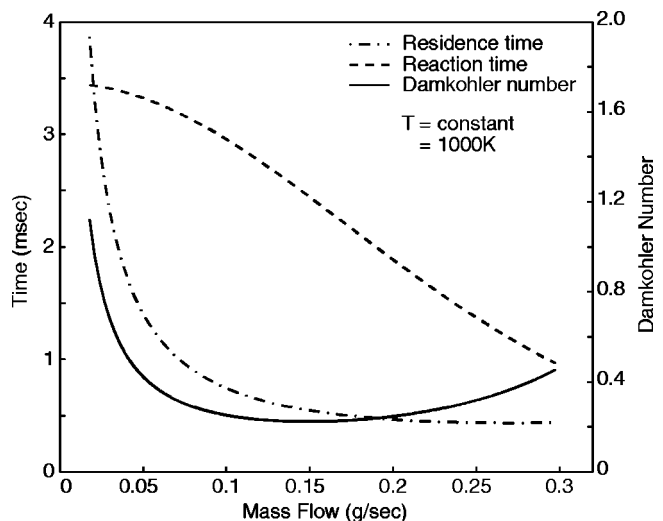


Fig. 9 Reaction time, residence time, and Da_h

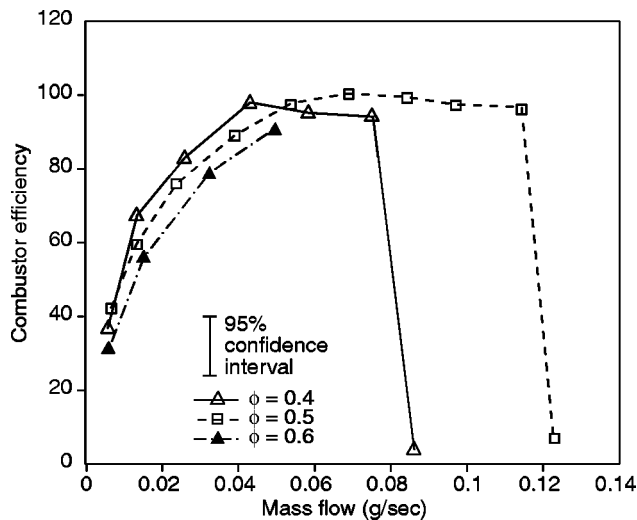


Fig. 10 Efficiency for slotted inlet six-wafer combustor

combustor, however, the slotted-inlet device produced higher exit temperatures and efficiencies in the high mass flow, kinetically limited regime of the operating line. This is attributed to the presence of multiple recirculation zones at the inlet which rapidly and uniformly ignite the incoming mixture. The sharp drop in performance for the slotted combustor is due to the rapid extinction of the ignition zones. In the annular case shown in Fig. 8, the single large recirculation zone stays lit over a wider range of mass flow rates, however, it also results in lower chemical efficiency.

Numerical simulations were performed on both inlet geometries using FLUENT 5 with a nine-species, 20-step, hydrogen-air reaction mechanism for the chemical model [8]. Figure 11(a) shows CFD generated contours for the combustor with the annular inlet. In Fig. 11(b) contours of temperature are plotted for a two-dimensional cut through the combustor inlet in the θ - z plane. These contours show a temperature gradient from the bottom of the combustion chamber to the top and indicate the primary ignition zone is in the upper right corner of the combustor. Figures 12(a) and 12(b) show the same contour plots for a combustor with a slotted inlet. The temperature contour in the θ - z plane shows significantly higher temperature in the lower regions of the combustion chamber and higher overall temperatures. The shape of

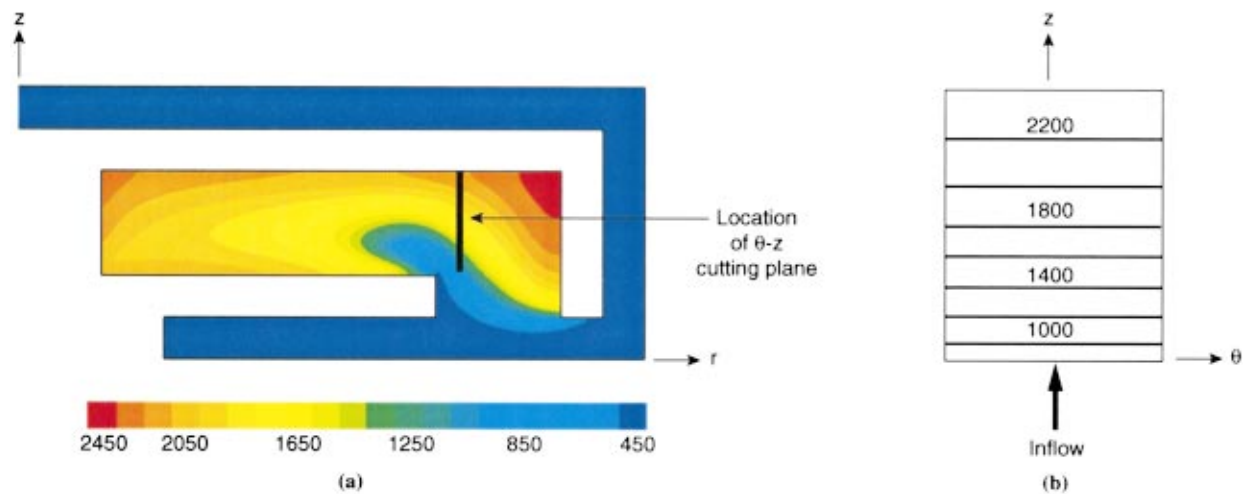


Fig. 11 Temperature (K) contours for annular inlet combustor

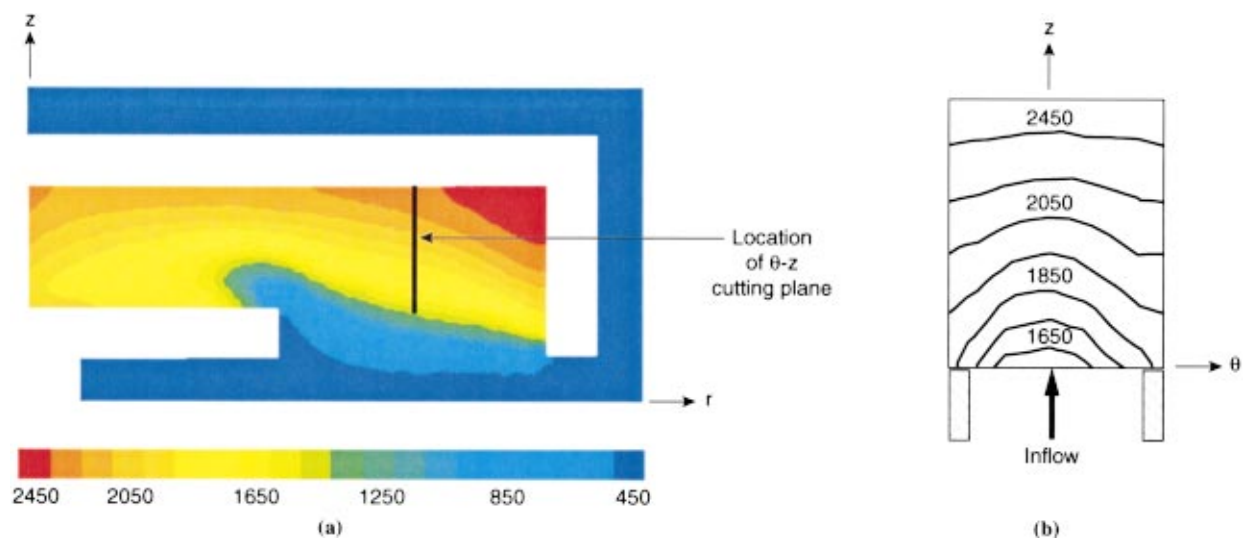


Fig. 12 Temperature (K) contours for slotted inlet combustor

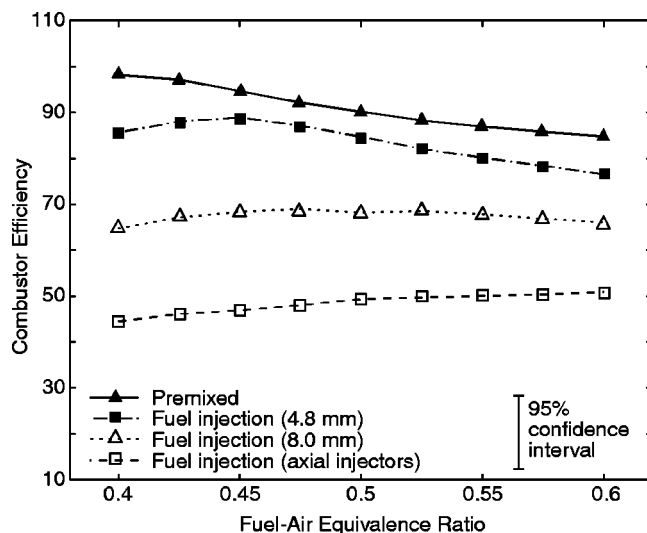


Fig. 13 Efficiency for fuel injection schemes

the contours indicates that there are small, hot ignition zones near the slotted inlet. Additional details of the numerical simulations can be found in reference [9].

4.3 Fuel Injection Schemes. The performance of the three sets of fuel injectors was evaluated by comparing efficiencies with the premixed case. Premixed operation consisted of fuel/air injection through the primary air inlet only. Figure 13 plots efficiency for different equivalence ratios at a mass flow rate of 0.045 g/s. The injectors were located at radii of 4.8 mm and 8.0 mm in the cooling jacket while a set of side-wall injectors were at the base of the cooling jacket. The figure indicates that performance decreases as the injectors get closer to the combustion chamber. This is due to the decreased mixing length.

4.4 Baseline Dual-Zone Hydrogen Tests. Baseline dual-zone testing was performed with hydrogen fuel injection through upstream fuel injectors at a radius of 8.0 mm. The dilution hole radial location was 7.0 mm. Plots of exit gas temperature and efficiency are shown in Fig. 14. This combustor achieved exit gas temperatures in excess of 1600 K at mass flow rates exceeding 0.12 g/s and efficiencies over 85%.

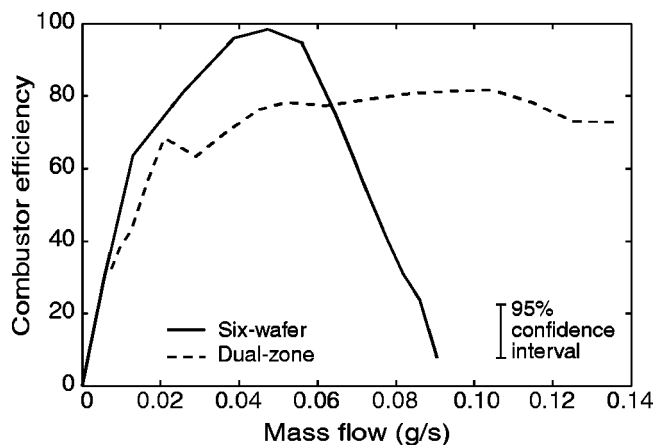
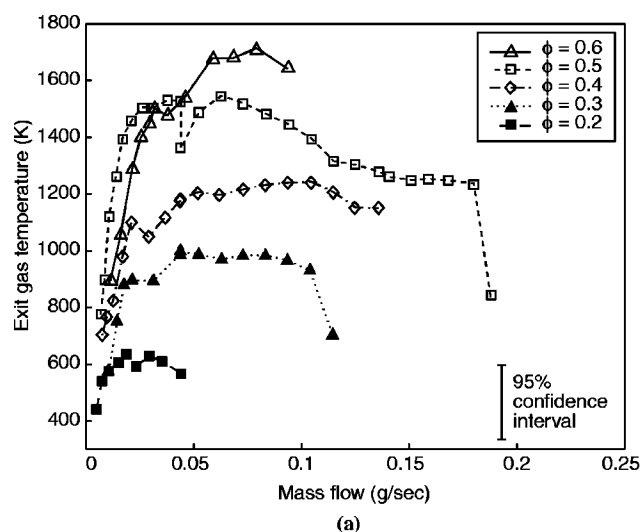


Fig. 15 Efficiency comparison of dual-zone combustor to baseline six-wafer combustor

The dual-zone combustor exhibited slightly lower exit gas temperature and efficiency than the single-zone, six-wafer combustor. However, its operating range was much wider. Combustion was stabilized at equivalence ratios as low as 0.2 and mass flow rates as high as 0.2 g/s. Figure 15 compares the dual-zone combustor efficiency to the single-zone for similar conditions.

The significant extension in operating range is largely due to strengthened recirculation zones and the hot, stable primary zone. For the same overall efficiency, the dual-zone combustor achieved a 100% increase in mass flow rate before blow-out. The ability to operate at low overall equivalence ratio is a result of a higher local equivalence ratio in the primary zone. It should be noted that for the current microengine applications, combustor efficiencies of 80% or higher are acceptable.

4.5 Effect of Primary Zone Size. To evaluate the effects of primary zone size, dilution holes located at radii of 5.9 mm and 7.0 mm were tested. These correspond to primary zone volumes of 78.5 mm³ (41% of total combustor volume) and 37 mm³ (20% of total combustor volume) respectively. Figure 16 plots efficiency for the two devices operating at similar conditions. Higher mass flows were achieved with the larger primary zone. This is largely due to increased residence time in the primary zone. There

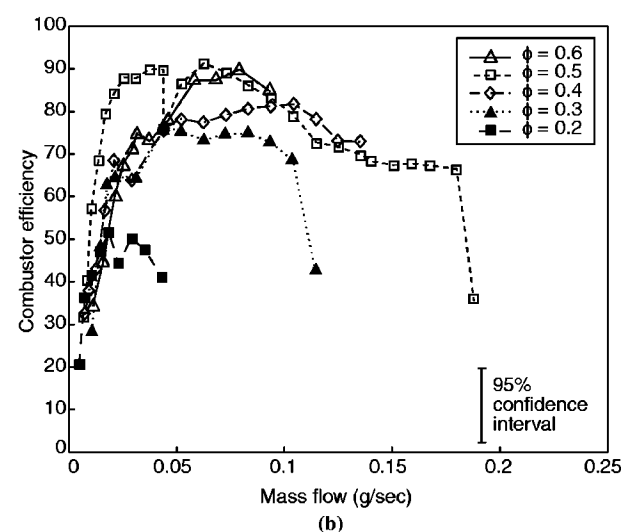


Fig. 14 Plots of exit gas temperature and efficiency for dual-zone combustor with hydrogen-air

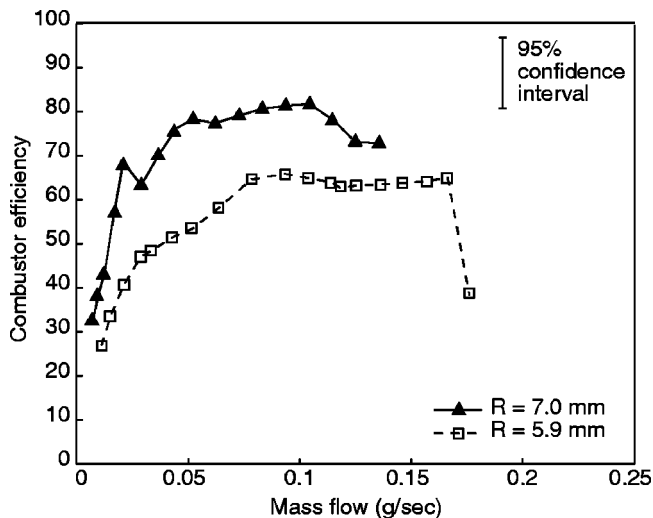
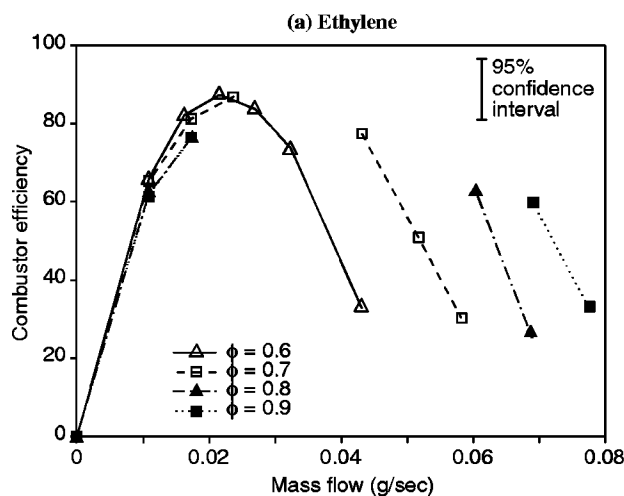


Fig. 16 Comparison of dual-zone combustors with different primary zone sizes

was also a reduction in efficiency. This was probably due to nonuniform exit temperature profiles where the measurement was made.

4.6 Hydrocarbon Tests. Hydrocarbon fuels with slower reaction rates were tested in the devices. For a micro engine to be practical, fuels such as ethylene (C_2H_4) and propane (C_3H_8) will be required due to their high energy densities and favorable storage properties. Results from combustion tests for these two fuels in the six-wafer combustor are shown in Fig. 17. The ethylene-air mixture achieved maximum power density at an equivalence ratio of 0.9 and 1 atm pressure, with exit temperature exceeding 1400 K and efficiency of 60%. The estimated power density for these conditions is $\sim 500 \text{ MW/m}^3$. The reaction rates of propane are significantly lower than ethylene and as a result, combustion could only be stabilized in the device at an equivalence ratio of 0.8. Exit temperatures of 1200 K were reached at 1 atm with an efficiency of 55%. This corresponds to a power density of $\sim 140 \text{ MW/m}^3$. Nominal residence times for ethylene-air and propane-air combustion in this device were 1.6 ms and 2.8 ms, respectively. Typical reaction times are 0.5–1 ms for ethylene-air and 1–2 ms for propane-air whereas hydrogen-air reaction times are approximately 0.2 ms.



The dual-zone combustor did not provide an expected broader range of mass flow for hydrocarbon-fueled operation. This is believed to be due primarily to inadequate mixing lengths, since the fuel injectors were designed for operation with hydrogen. As a result, mass flows of only up to 0.06 g/s and efficiencies less than 50% were achieved with ethylene. However, these combustors were operated with a wider range of equivalence ratios. This was due in part to fuel injection, which prevented upstream burning and allowed higher mixture ratios. In addition, lower equivalence ratios were possible due to the strengthened recirculation zones created by the dilution flow. Figure 18 shows efficiency for both ethylene-air and propane-air combustion in a dual-zone combustor with dilution holes located 7.0 mm radially outward. The maximum power density achieved for the ethylene-air mixture was $\sim 100 \text{ MW/m}^3$ and 75 MW/m^3 for propane-air with residence times of 2.1 ms and 3.7 ms, respectively.

5 Operating Space

Figure 19 plots the empirically identified operating space for the baseline six-wafer and the dual-zone microcombustors burning hydrogen. The operating boundaries include:

1. the lean blowout limit,
2. the upstream burning limit where combustion in the cooling jacket occurred,
3. the structural limit, and
4. the 1600 K T_{t4} micro-engine cycle limit (assuming 80% combustor efficiency).

A narrow operating corridor exists between the lean blowout limit and the upstream burning limit, however, as previously discussed, the design mass flow rate of 0.36 g/s was not reached. As a result, initial engine designs will have reduced flow (225 μm compressor blade height instead of 400 μm blade height).

The dual-zone combustor has a broader operating space due to the stable, hot primary zone. The most notable features of this operating space are: the lower lean blowout limit and the absence of an upstream burning limit due to non-premixed operation.

Figure 20 shows data for all of the combustors, including the three-wafer device mentioned in Section 2.6, presented in a plot of Damköhler number versus chemical efficiency. The wall temperature measurements along with a simple one-dimensional heat transfer model were used to estimate the heat loss from the devices enabling a calculation of chemical efficiency, [5]. The residence time was calculated using the exit gas temperature measurements while Chemkin III, a reaction kinetics software package, was used to compute chemical time-scales. The nine-species, 20-

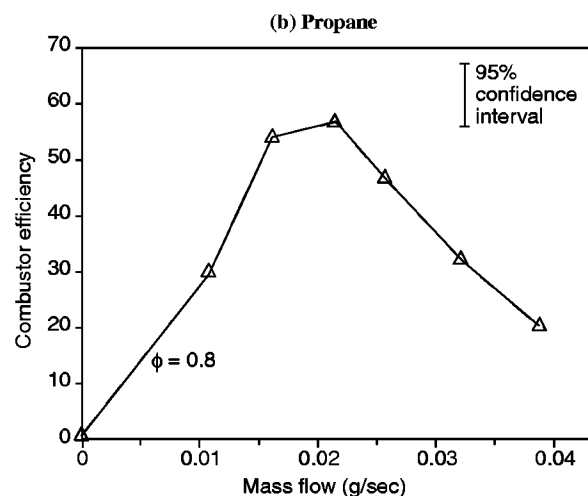


Fig. 17 Efficiency for ethylene and propane in the baseline six-wafer combustor

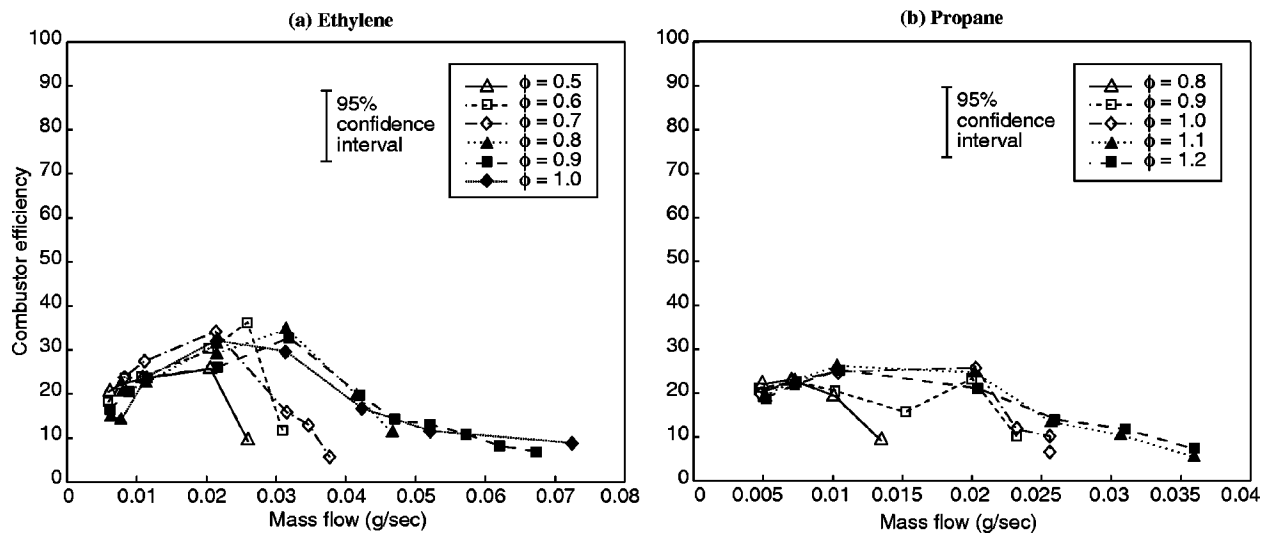


Fig. 18 Efficiency for ethylene and propane in the dual-zone combustor

step hydrogen-air reaction mechanism used for the CFD simulations was utilized as the chemical model, [8]. For hydrocarbon-air cases GRIMech 3.0 was used. As Damköhler number increases, chemical efficiency asymptotically approaches unity. Points to the far left with low chemical efficiency are a result of the combustor approaching blowout in the high mass flow regions of the operating space.

Although Fig. 20 shows high chemical efficiency over a range of Damköhler numbers, the performance of these devices was found to be a function of both chemical and thermal losses. In this case performance is quantified in terms of nondimensional power density, or actual power density normalized by the maximum possible power density at those conditions.

If it is assumed that there is adequate fuel-air mixing and rapid, uniform ignition, this parameter is a function of only two parameters; Damköhler number and nondimensional heat loss. Damköhler number was estimated as previously described using exit gas temperature measurements for residence time calculations and Chemkin III for chemical time approximations. Heat loss can be cast in nondimensional terms by normalizing the actual heat lost from the device by the maximum possible heat which can be

generated at a given operating condition. Again, the actual heat lost was estimated using a one-dimensional heat transfer model, [5].

Figure 21 plots this nondimensional power density for all of the combustors in three-parameter space. The plot includes all three fuels tested (hydrogen, ethylene, and propane). The heat loss parameter and Damköhler number are on the x and y-axes, respectively. A second-order least-squares surface fit was used to generate contours of nondimensional power density (the fit explains 62% of the variance in the data).

Thermal efficiency in these devices is inversely proportional to heat loss while chemical efficiency is directly proportional to Damköhler number. Hence, optimum performance is achieved at low levels of heat loss and high Damköhler number as indicated in Fig. 21.

This nondimensional operating space further explains the operating line considerations of a microcombustor. Conditions of increasing mass flow for constant equivalence ratio move from the upper right side of the operating space to the lower left. In the low mass flow region of the space, performance is limited by heat loss, while at high mass flow Damköhler number falls and performance

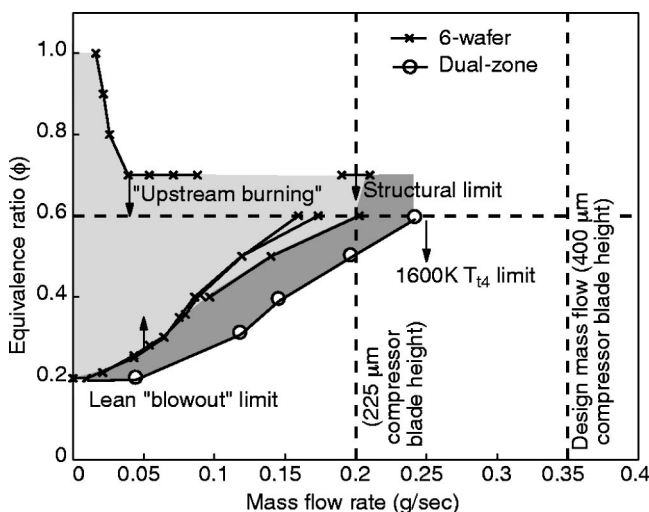


Fig. 19 Operating space for six-wafer and dual-zone combustors

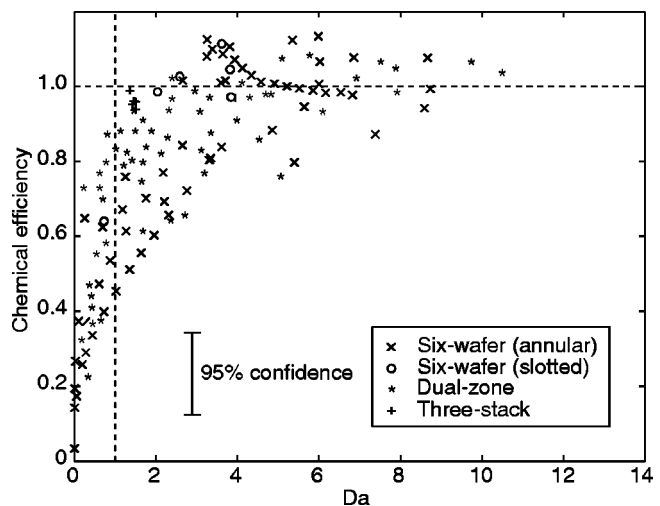


Fig. 20 Damköhler number versus chemical efficiency for several microcombustors

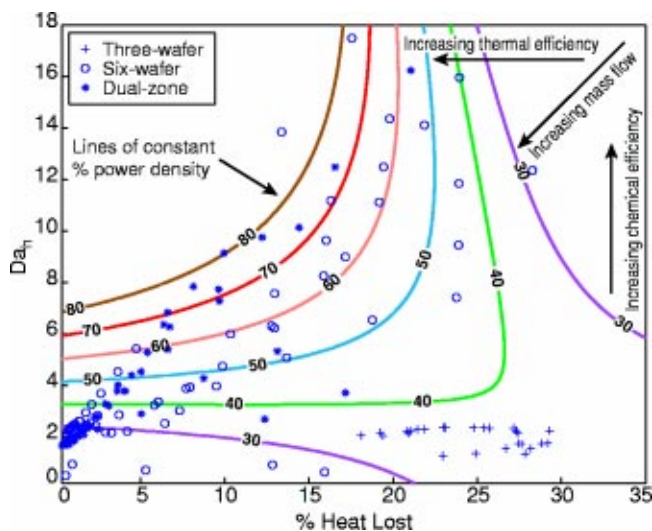


Fig. 21 Nondimensional operating space

is limited by chemical inefficiency. In the area of moderate mass flow, the center of the operating space, the highest power densities are achieved. This is consistent with the data presented in Section 4 where the highest overall efficiencies were found to be at moderate mass flow rates. Although we do not present the results here, the overall trends shown in Fig. 21 are consistent with what has been obtained from a PSR model with heat loss run for the same operating space.

Figure 21 also indicates that the microcombustor design changes discussed previously improved overall performance. The three-wafer combustor suffered from high heat loss and lies to the lower right in the operating space. The largest performance enhancement came with the implementation of the cooling jacket for thermal isolation of the combustion chamber. The six-wafer data lies much farther to the left (low heat loss region) and also has higher Damköhler numbers. The dual-zone combustors provided a hot stable flame-zone which resulted in higher Damköhler numbers and improved performance.

This nondimensional operating space can also be useful in the design process. For example, a good estimate of the volume required for a propane-air microcombustor suitable for a micro gas turbine can be obtained. If similar conditions to those run here are assumed (1600 K exit gas temperature, 1 atm pressure, and mass flow of 0.15 g/s) and an equivalence ratio of $\phi=0.8$ is selected, the nondimensional power density can be computed and is approximately 60%. The heat loss from the device can be assumed to be approximately 5% based on previous data with thermal management techniques such as the cooling jacket. From Fig. 21, the desired Damköhler number can be found and is approximately 5.5. The chemical time can be computed using Chemkin III or any other suitable technique and the desired residence time found from the Damköhler number. This results in a residence time of 1.44 ms and a volume of 950 mm³, or about a fivefold increase in volume compared to those combustors presented here.

6 Design Recommendations

The following design trades and recommendations for micro-combustion systems have been identified:

1. the high power density requirements of a micro-combustor mandate high mass flow rates per unit volume. However, this results in low residence time and Damköhler numbers of order unity. Microcombustion systems are fundamentally limited by this tradeoff between high power density and efficient combustion.
2. the large surface-area-to-volume ratio of a microcombustor makes nonadiabatic operation intrinsic to its design. The coupling

between heat transfer and chemical kinetics is a critical element of the design process. Thermal isolation of the device improves thermal efficiency, however, cold walls are detrimental to reaction rates. Strategies such as the cooling jacket presented here are effective solutions. While keeping the inner wall cool, it acts as a fuel-air pre-heater to improve kinetic rates, offsetting the effect of lower wall temperatures.

3. recirculation zones are critical for rapid and uniform ignition of incoming reactants. However, if these ignition zones are too large, the effective volume of the combustion chamber will be significantly reduced resulting in decreased residence time and Da_h . Flame holding ignition zones should be carefully tailored to balance these two effects.

4. fuel injection for non-premixed operation should occur as far upstream of the combustion chamber as possible to provide adequate mixing length.

5. it is not sufficient to design a microcombustor exclusively for design point operation. Operating line issues need to be considered. For the combustors presented here, the minimum Da_h occurs at mass flows and pressures corresponding to operation prior to choking of the nozzle guide vanes.

7 Summary and Conclusions

Although microscale combustion systems have similar functional requirements to conventional combustors, residence time and heat transfer effects significantly impact the design. To explore these phenomena at the microscale, two devices were fabricated and tested while numerical simulations were also conducted.

The test devices consisted of six silicon wafer levels and a combustion chamber measuring 191 mm³. Premixed hydrogen-air combustion was sustained at exit temperatures greater than 1600 K, efficiencies over 90%, and power densities up to 1100 MW/m³. Non-premixed hydrogen-air combustion was stabilized with similar performance. Hydrocarbon fuels were also burned in the devices and achieved efficiencies of up to 60% and power densities near 500 MW/m³. It was found that the low mass flow regime of the operating space was limited by thermal losses, while in the high mass flow regime, chemical inefficiency dominated. Multiple fuel injection schemes and inlet geometries were tested. The importance of recirculation zones to rapidly and uniformly ignite reactants was identified and supported by numerical simulations and analysis. The operating space for these combustors was mapped and additional boundaries such as upstream burning were noted. The broadening of this space by the implementation of a hot primary zone and dilution region was observed. A non-dimensional operating space was developed and shown to be useful as a design tool.

Experimental and computational work is continuing with a focus on a hydrocarbon-fueled microcombustor. Alternate combustion strategies are being investigated and a catalytic hydrocarbon combustor is in development.

Acknowledgments

This work was conducted as part of the MIT Microengine Project supervised by Prof. Alan Epstein. The authors are grateful to Professor Epstein for his direction throughout the course of this research and also to Prof. Martin Schmidt of the MIT Microsystems Technology Laboratory and Dr. A Ayon. We also acknowledge the technical support of Greg Simpson at Thunderline-Z Inc. and the Gas Turbine Laboratory technical staff; Viktor Dubrowski, James LeTendre, and William Ames. The administrative support of Lori Martinez and Holly Anderson, as well as the editorial assistance of Diana Park, is also greatly appreciated. The authors also thank the entire microengine team especially Dr. Eugene Huang, Dr. Chris Cadou, Jongwoo Peck, and Khoon Tee Tan. All devices were fabricated at the MIT Microsystems Technology Laboratory; the assistance of the technical staff is also acknowledged. This work has been sponsored by the Army Research Of-

fic (DAAG55-98-1-0292) under Dr. T. Doligalski and by DAPRA (DAAG55-98-1-0365, DABT63-98-C-0004) under Dr. R. Nowack and Dr. R. Rosenfeld, respectively.

Nomenclature

A	= Arrhenius pre-exponential factor
a, b	= Arrhenius exponents
d_h	= hydraulic diameter
Da_h	= homogeneous Damköhler number
E''	= heat loss
E	= heat generated
E_a	= activation energy
h	= enthalpy
LHV	= lower heating value
$\dot{m}$	= mass flow
P	= pressure
Q_{loss}	= heat loss from combustor
R	= gas constant
T	= temperature
V	= volume
η_c	= overall combustor efficiency
η_{chemical}	= chemical efficiency
η_{thermal}	= thermal efficiency
ρ	= fluid density
τ_{reaction}	= reaction time
$\tau_{\text{residence}}$	= residence time

Subscripts

a	= air
f	= fuel
0	= initial condition
1	= combustor inlet
2	= combustor exit

References

- [1] Epstein, et al., 1997, "Micro-Heat Engines, Gas Turbines, and Rocket Engines," presented at the 28th AIAA Fluid Dynamics Conference.
- [2] Groshenry, C., 1995, "Preliminary Study of a Micro-Gas Turbine Engine," S. M. thesis, Massachusetts Institute of Technology, Cambridge, MA.
- [3] Waitz, I. A., Gautam, G., Tzeng, Y.-S., 1998, "Combustors for Micro Gas Turbine Engines," ASME J. Fluids Eng., **20**, pp. 109–117.
- [4] Mehra, A., and Waitz, I. A., 1998, "Development of a Hydrogen Combustor for a Microfabricated Gas Turbine Engine," presented at the Solid-State Sensor and Actuator Workshop at Hilton Head, SC.
- [5] Mehra, A., 2000, "Development of a High Power Density Combustion System for a Silicon Micro Gas Turbine Engine," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.
- [6] Mehra et al., 2000, "A 6-Wafer Combustion System for a Silicon Micro Gas Turbine Engine," J. Microelectromech. Syst. .
- [7] Harrison, T., 2000, "Packaging of the MIT Microengine," S.M. thesis, Massachusetts Institute of Technology, Cambridge, MA.
- [8] Yetter, R. A., Dryer, F. L., and Rabitz, H., 1991, "A Comprehensive Reaction Mechanism for Carbon Monoxide/Hydrogen/Oxygen Kinetics," Combust. Sci. Technol., **79**, pp. 97–128.
- [9] Lee, J., 2000, "Numerical Simulation of a Hydrogen Micro-combustor," S.M. thesis, Massachusetts Institute of Technology, Cambridge, MA.

Two-Dimensional Flow and NO_x Emissions in Deflagrative Internal Combustion Wave Rotor Configurations

K. Pekkan

Post Doctoral Fellow

M. R. Nalim

Assistant Professor

Mechanical Engineering Department,
Purdue School of Engineering and Technology,
Indiana University—Purdue University
Indianapolis,
723 West Michigan Street,
Indianapolis, IN 46202

A wave rotor is proposed for use as a constant volume combustor. A novel design feature is investigated as a remedy for hot gas leakage, premature ignition, and pollutant emissions that are possible in this class of unsteady machines. The base geometry involves fuel injection partitions that allow stratification of fuel/oxidizer mixtures in the wave rotor channel radially, enabling pilot ignition of overall lean mixture for low NO_x combustion. In this study, available turbulent combustion models are applied to simulate approximately constant volume combustion of propane and resulting transient compressible flow. Thermal NO production histories are predicted by simulations of the STAR-CD code. Passage inlet/outlet/wall boundary conditions are time-dependent, enabling the representation of a typical deflagrative internal combustor wave rotor cycle. Some practical design improvements are anticipated from the computational results. For a large number of derivative design configurations, fuel burn rate, two-dimensional flow and emission levels are evaluated. The sensitivity of channel combustion to initial turbulence levels is evaluated. [DOI: 10.1115/1.1586315]

Introduction

A wave rotor machine utilizes nonsteady wave motion to exchange energy by direct work action between chemically reacting or inert fluids. It consists of a large number of straight channels arranged about the axis of a rotor. By rotation the channel ends are periodically ported to connected steady-flow components. The rotation speed of the rotor is low compared to conventional turbomachines. For detailed descriptions of wave-rotor principles and applications see [1,2]. In this study it is assumed that an internal combustion wave rotor (ICWR) is coupled to a compressor and turbine, which are placed at its inlet and outlet, respectively.

Previous simulations of wave rotor processes have provided good overall indications of wave rotor behavior and performance potential. One-dimensional simulations modeled most loss mechanisms, heat transfer, and combustion in a simplified manner, [3,4,5]. Two-dimensional laminar computations included duct flows, and identified major flow features that influence performance and thermal loads, [6]. An inviscid three-dimensional study with rotation can be found in [7]. Feasible modes of combustion wave rotors have been assessed and one-dimensional simulations of basic cycle designs have been performed, [8,9]. Ignition and flame propagation in a wave-rotor-like channel has been experimentally demonstrated, [10]. These studies indicate that it is feasible to design combustion wave rotors with reasonable size and fuel-rate for most gas turbine applications, resulting in pressure gain with lower weight and simpler integration than pressure-exchange wave rotors.

Two major questions that must now be addressed are pollutant emissions of such a system, and the thermal loads on the rotor and other components. Numerical studies of a stratified-charge combustion wave rotor indicate that hot gas leakage and thermal gradients/loads may be a significant problem, [9]. For the pressure-exchange wave-rotor cycle designed for gas turbine top-

ping, the one-dimensional and two-dimensional estimates of wall and exit temperatures apparently disagree, creating uncertainty about thermal loads, [6,11]. In general, among remaining challenges affecting the performance of wave rotor machines are: leakage, internal heat transfer, hot gas and air mixing in channels, loss of total pressure at inlet and exit openings, and complex ducting.

One-dimensional gas dynamics codes developed for wave rotor calculations are popular and practical for cycle design, [5], preliminary analysis, and system/component matching, [12]. They are also used to determine the relative importance and sensitivity responses of various design parameters. Both off-design, [13,14], and periodic steady-state studies can be performed with these codes, and key physical phenomena can be modeled with satisfactory accuracy, except for heat transfer and combustion. Models exist for viscous losses, partial opening, leakage, [11], and stratified combustion, [9]. Nevertheless, multi-dimensional simulations are now needed to provide essential flow details and offer a necessary tool for the analysis of some new design ideas. Such simulations are still expensive, partly because transient codes are required. Requirements for refined meshes near turbulent boundary layers and traveling discontinuities (shock and expansion waves, detonations) results in very small time steps and considerable computer time to obtain a complete cycle simulation. Therefore, this study provides two-dimensional flow analysis, focusing on different and improved combustion models, to elaborate and verify previous one-dimensional studies and lead to a simple two-dimensional cycle design approach.

More specifically, this study is aimed at investigating a special feature proposed for combustion wave rotors, which necessitates multi-dimensional analysis. A simple method of pilot ignition aids initiation of a deflagrative or detonative combustion process while allowing for lean-burn, and simultaneously reducing the leakage of hot combustion gases and thermal load on the internal surfaces of a combustion wave rotor. The method consists of creating precise circumferential partition walls in the inlet duct and, as importantly, in the wave rotor channel itself. Note that the use of radial partition walls in the inlet duct has been previously described, [9]. Such partition walls provided circumferential varia-

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Amsterdam, The Netherlands, June 3–6, 2002; Paper No. 2002-GT-30085. Manuscript received by IGTI, December 2001, final revision, March 2002. Associate Editor: E. Benvenuti.

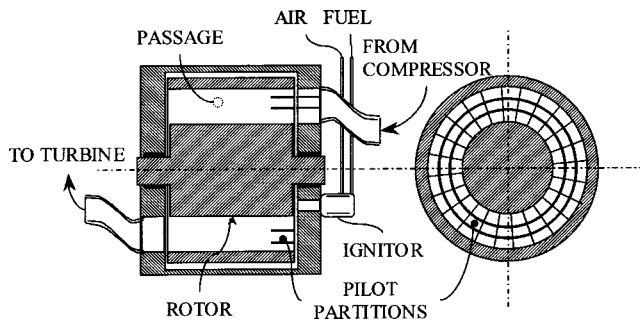


Fig. 1 Internal combustion wave rotor (ICWR) with pilot partition walls

tion in the fuel distribution in the inlet, and allow a longitudinally stratified fuel-air mixture to be introduced into the wave rotor channels. This provides ignitable mixtures near the ignition location, and combustible conditions in the remaining mixture.

The use of circumferential partition walls in the current study provides stratification in the radial direction, adding a new level of control of the combustion process. In combination, the partition walls provide a limited region of the channel to be charged with a near-stoichiometric or enriched mixture that ensures ignition or initiation of a deflagrative or detonative combustion process. Similar to pilot combustion in various combustion engines, the enrichment may be with fuel, oxidant, or both. The combustion in this region is controlled by the provision of circumferential partition walls in the rotor channels that extend a relatively short distance into the length of the channel. Figure 1 shows a wave rotor with circumferential partitions near the inlet. The inlet duct partitions can be both circumferential and radial, with different fuel distribution schemes in selected inlet zones.

The significance of the new partitions is multifold; first, it provides for a pilot ignition zone that can be charged by an enriched mixture and separately ignited. This allows for the remaining mixture to be made significantly leaner in fuel than otherwise possible and therefore lower in nitrogen oxide (NO_x) emissions. Second, it provides a noncombustible gas to be placed adjacent to the hub and shroud, thus eliminating the leakage of hot gas from the inlet side of the rotor, and helping to cool the rotor. This will minimize thermal damage of the bearings and other components of the wave rotor, and provide some control of the rotor temperature and its gradient. Third, the most readily ignited mixture is supplied only to the central section, and it avoids pre-ignition by hot gas leaking into the low-pressure inflowing mixture. Fourth, there may be structural advantages as envisaged by Kentfield [15], but this effect is not studied here.

1 The Deflagrative Cycle

In Fig. 2, a developed view of a typical ICWR is shown with stratified charging, radial inlet partition walls and dual-sided deflagrative combustion. In Table 1 the given valve timing sequence considers both radial and circumferential partitioning. Phases I, II, and III are analyzed in this study.

2 Computational Model

A two-dimensional model is developed to compute the transient flow and combustion efficiently, in a channel with circumferential partition walls. For rotors with narrow channels, sidewall effects may not be negligible, but require a three-dimensional time-consuming computation.

2.1 Passage Geometry and Grid. Dimensions of the wave rotor channel that is being studied are shown in Fig. 3, which are representative of existing test wave rotor designs published in the literature, Table 2.

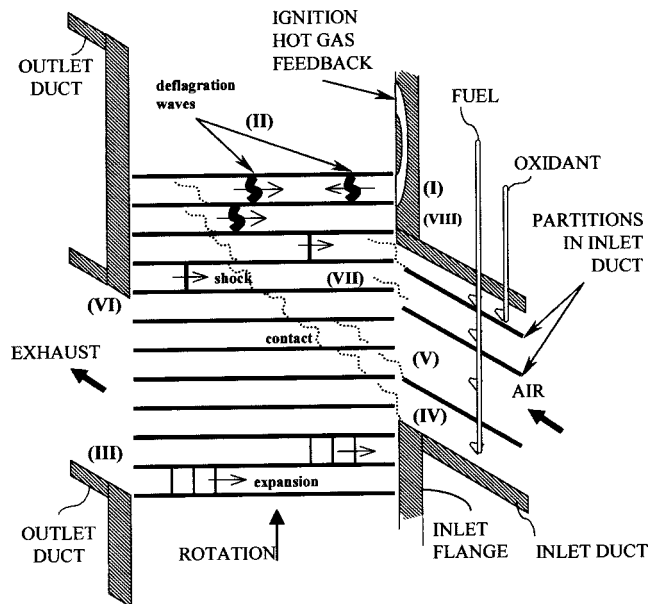


Fig. 2 Internal combustor wave rotor with longitudinal stratification: Developed ("unrolled") view and basic wave dynamics. Refer to Table 1 for phase number descriptions.

The total number of computational cells, including the solid cells of partition walls is 11,968, Fig. 4. The upper limit of grid aspect ratio is set to 9. Grid sensitivity tests showed no phase errors and differences are smaller than the variations, evaluated in this study, due to model assumptions.

2.2 Solver. Calculations are performed using the STAR-CD code, a general purpose, finite volume algorithm that operates on a versatile unstructured grid system to solve the conservation equations. Its modeling capability includes compressible, multi-phase, and chemical reacting flows, through a choice of different Reynolds-averaged turbulent models, [16]. Extended versions of the SIMPLE [17] and PISO [18] algorithms are used for steady-state and transient calculations. Spatial differencing is second order, and a fully implicit first-order temporal differencing scheme is used.

For the presented results, normalized solver tolerances for all variables were set to 0.01, except 0.0001 for pressure. Convective terms for scalar equations are discretized with a blended upstream

Table 1 Proposed deflagration cycle

Phase	End (ms)	Description
I	0.065	hot gas injection (ignition)
II	3.865	constant volume combustion
III	4.585	exhaust port opens to turbine
IV	4.615	A little cold air is introduced from the pilot partition inlet to avoid pre-ignition. Lean fuel-air mixture is introduced from the remaining divisions of the inlet port (above and below the pilot partition).
V	4.815	Turbine exhaust continues. Lean mixture is introduced from all divisions of the inlet port.
VI	4.995	Exhaust port closed, shock-compression starts. Lean mixture inflow continues.
VII	5.055	Shock compression process continues. Richer mixture is introduced through the pilot partition inlet. Lean mixture inflows from the remaining divisions.
VIII	5.245	The richer pilot mixture continues to enter to the pilot partition. Cold-leakage control air is introduced into the remaining partitions. The initial mixture distribution is finally obtained.

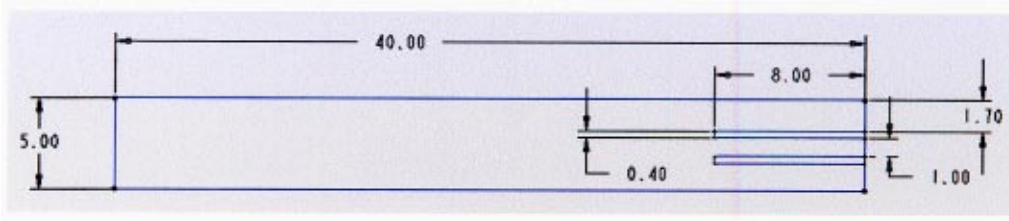


Fig. 3 Model dimensions (in cm)

and second order monotone advection and reconstruction scheme (MARS), for all other equations MARS scheme is employed. The total enthalpy form of the energy equation is solved with chemicothermal terms for scalars. For temperature and scalars a relaxation factor of 0.95 is used. The maximum Courant number was five times less than the solver allowance.

2.3 Turbulence. Intense turbulence is achievable and expected during the combustion period of a wave rotor, due to the immediately preceding high-speed inflow. The standard high Reynolds number k - ϵ model with algebraic law-of-the-wall functions is employed. This model has been used successfully in reactive flow studies with mixing and kinetics controlled conversion rate models, especially in internal combustion (IC) engines, [19,20]. It is valid only in the region of turbulent flow, where the local turbulent Re number ($\rho k^2/\mu\epsilon$) is high. Despite the usual uncertainties involved in turbulent flows, useful results are being obtained, [21]. Model details can be found in [22,23]. Terms for compressibility and wall roughness effects are also present in the turbulence model used. Values assigned to model coefficients are given

Table 2 Selected internal combustor wave rotor specifications

	Nalim [4]	Bilgin [10]	Kentfield [15]	Nalim [29]	Current Study
Mean radius (cm)	8.15	N/A	9.53	25.8	N/A
Rotor length (cm)	15.24	40.6	25.4	60.6	40
Channel height (cm)	2.18	4.0	6.35	9.09	5.0
Cycles/second	560	N/A	100	191	190.7
Inlet area (cm ²)	26	16	10.6	-	N/A

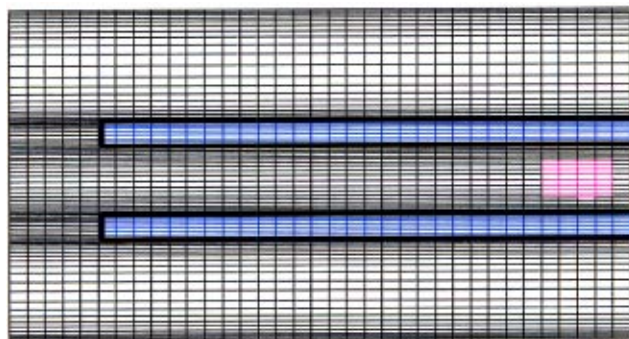


Fig. 4 Computational grid in the partitioned region (closeup)

Table 3 Values assigned to standard k - ϵ turbulence model coefficients

	C_μ	σ_k	σ_ϵ	σ_h	σ_m	$C_{\epsilon 1}$	$C_{\epsilon 2}$	$C_{\epsilon 3}$	σ_{fu}	σ_f	$C_{\epsilon 4}$	K	E
Hjertager [24,25]	0.09	1.0	1.3	0.7	-	1.44	1.92	-	0.7	0.7	-	-	-
Hessel [46]	0.09	1.0	1.3	-	-	1.44	1.92	-1.0	-	-	-	-	-
Current study	0.09	1.0	1.22	0.9	0.9	1.44	1.92	0.0 1.44	0.9	0.9	0.33	0.42	9.0

in Table 3. These values are similar to [24], where for cylindrical flames $C_{\epsilon 2}$ is 1.79. The only notable change is in the turbulent Schmidt values of mixture fraction and mass fraction of fuel where slightly higher values are used compared to [24] and [25].

For k - ϵ models it is known that the prediction of swirling flows and recirculation zones is poor [20,26], compared to Reynolds stress equation models. Reynolds stress models (RSE) or higher-order closures can be used for simple geometries, however, in practical combustion devices less sophisticated representations have to be used because of computer capability limitations. For turbulent jet diffusion flames, [20], the superiority of RSE predictions is reported, together with the poor performance of the low Reynolds number models for the test case studied.

The turbulent time scale is in the order of one milliseconds and turbulence decays relatively quickly in the motionless channel regions. Turbulence that sustains combustion is generated by the jet-like partition exit flow. Thus the severity of the initial turbulence level in the fuel partition is critical for a high kinetic energy partition exit flow and subsequent high combustion rates.

For continuous flow devices, initial k and ϵ values can be estimated by making assumptions for turbulence intensity, T_i and characteristic length scale, l , [27]. However, for a charged wave rotor state with the inlet valve just closed, there is no internal mean velocity. For this reason, empirical estimates from spark ignited (SI) engine studies are employed. In these machines, the intake process generates a highly turbulent flow ($1.4 > u'/\bar{U} > 0.2$) that may persist throughout the compression phase, [28]. In the intake process of wave rotor engines the value of $u'/\bar{U}$ is probably smaller than that of the SI engines; however, intake $\bar{U}$ is high and the subsequent shock-compression process will intensify turbulence.

In Table 4, selected approaches for the estimation of initial turbulence level are summarized. Using these approaches, a total of eight different runs are performed with varying initial turbulent kinetic energy and dissipation rates, Table 5.

2.4 Boundary Conditions. The symmetry sidewall boundary conditions simulate a wave rotor channel with infinitely extending parallel hub and shroud walls. Much of the combustion takes place at constant volume, with inlet and outlet ends specified as solid walls. The turbine exhaust and subsequent mixture injection processes are specified by the use of pressure boundary conditions at inlet and outlet. Concentrations at outlet boundaries are

Table 4 Initial turbulence level estimation via different relations. ($\tau_e = l/u'$ is eddy decay time, k_o and ε_o are the initial values.)

Estimate Source	Key Relations	k_o (m ² /s ²)	ε_o (m ² /s ³)	τ_e (ms)
Versteeg [27], general internal flows	$k = K \cdot (\overline{UT_i})^2$ $\varepsilon = C_\mu^{3/4} \frac{k^{3/2}}{l}$ $K = 0.5 \sim 1.5$ $l = (0.07 \sim 0.1)L$	100 ($\overline{U_i} = 100$ m/s) ($T_i = 0.1$)	$0.23e^6$ ($l = 0.7$ mm)	0.07
Tabaczynski [28], spark ignition engines	INTAKE: $u' \approx \overline{U}$ and $\varepsilon \approx (u')^3/l$ $l = 2 \sim 10$ mm TDC: $l \approx h$ h , chamber height	3600 ($\overline{U_i} = 60$ m/s)	$7.10e^6$ ($l = 5$ mm)	0.7 at IVC
Abraham [19], premixed charge engine combustion	$k = \alpha(\eta_v \cdot \text{rpm})^2$ $\varepsilon = \beta(\eta_v \cdot \text{rpm})^3$ $\alpha = 0.608$ $\beta = 0.305$ $\eta_v = 40 \sim 60\%$ 1000~1500 rpm	$2.4e^5$	$7.4e^7$	
Hessel [46], diesel engines at IVC	$k = C_k \overline{U}^2$ $\varepsilon = C_\varepsilon (\overline{U}^3 / \sqrt{A})$ $C_k = 0.165$ $C_\varepsilon = 0.055$	590	$2.4e^6$	
Kuo [40], port fuel injection engine at BTDC firing	$k = 0.9 \cdot \overline{U}_{\text{piston}}^2$ l is proportional to the closest wall distance	$1.5e^4$ ($\overline{U_i} = 130$ m/s)	$13.08e^7$	

first-order extrapolated. Hot gas injection is modeled by an inlet boundary, where the injection pressure is calculated from the specified values of density and temperature.

2.5 Turbulent Combustion Model. Combustion calculations implemented in quasi-steady one-dimensional wave rotor codes are quite developed, [9]. They are applied to practical test cases operating in different combustion modes, [8,29]. Multiple species or a single reaction progress variable is utilized to allow for a variable fuel-air ratio. Mixing controlled reaction is combined with a simple eddy diffusivity model. Other notable features that are incorporated are limit and combustion temperature factors and a simple total-energy based flammability limit, [9].

The turbulent combustion process in wave rotors resembles that of the internal combustion engines. For this reason models developed for SI and CI engines are believed to be applicable to wave rotor combustion phenomena. The conventional approach in modeling turbulent premixed combustion in SI engines is to use detailed chemistry or one-step irreversible Arrhenius kinetics for chemical combustion, combined with a mixing controlled conversion rate model, [19].

A brief review of the computational modeling alternatives are given in [26,30]. Among them, the presumed PDF and eddy breakup-type methods are two popular approaches for the closure

of mean reaction rate. Both methods with sub-variants are being applied to a large range of industrial combustion problems [31]. For gas turbine combustion with sprays, the eddy breakup model has provided some useful results, [32]. Both methods, applied to a steady bluff-body stabilized burner resulted identical solutions irrespective of the turbulence model used, [20]. In [33], a modified form of the eddy breakup reaction rate equation is used to predict premixed flame propagation and damaging overpressures in large (2.5 m dia.) and long cylindrical vessels. Engine simulations in which the reaction rates are predicted by mixing-controlled combined models are numerous, [32,34,35].

In this study, turbulent combustion of propane (C_3H_8) is modeled via the eddy-dissipation model of Magnussen [24]. The model is applicable to both premixed and non-premixed systems. Other closely associated models are the transport and algebraic eddy break up models of Spalding [36]. The reaction time scale T_R is calculated as the sum of turbulent kinetic, k/ε and chemical kinetic, $\rho m_{lr}/R_{kin}$ time scales. This combined approach prevents over-prediction of reaction rates near wall regions. The fuel reaction rate then follows from Eq. (1).

$$R_F = \frac{\rho}{T_R} A_{ebu} \cdot \min \left[m_F, \frac{m_O}{s_O}, B_{ebu} \frac{m_P}{s_P} \right] \quad (1)$$

Table 5 Turbulent test case configurations. All cases have pilot partitions.

CASE	Equiv. Rat.	k_o m ² /s ²	ε_o m ² /s ³	T_{tb} (s) = k_o / ε_o
1b (base)	0.8	3600	$7.1e^6$	$0.507e^{-3}$
1bt1	0.8	40.5	23460	$1.720e^{-3}$
1bt2	0.8	3.461	4584	$0.760e^{-3}$
1bt3	0.8	$1.5e^4$	$13.08e^7$	$0.110e^{-3}$
1bt4	0.8	3600	$0.89e^6$	$4.045e^{-3}$
1bt5	0.8	100	$0.23e^6$	$0.434e^{-3}$
1bt6	0.8	$2.4e^5$	$7.4e^7$	$3.240e^{-3}$
1bt7	0.8	590	$2.4e^6$	$0.246e^{-3}$

The eddy breakup coefficients A_{ebu} and B_{ebu} are used to control strength of the source. The absolute value of A_{ebu} can range between 2 to 10; the lower value is used if the reaction rate is controlled by the consumption of a reactant, and the higher, if by the rate of formation of a product. In this study 4 and 0.5 is used for A_{ebu} and B_{ebu} , respectively, which are the typical values used in the literature, [24,25]. In [37], a numerical method is proposed for the calculation of A_{ebu} explicitly as a function of the mean and standard deviation of mixture fraction PDF. Commonly accepted values may be in error up to two orders of magnitude, especially in the rich regions of the flame. Sensitivity studies for these constants need to be performed. The eddy breakup model with the

Table 6 Rate constants ($K=A.exp(B/T)$) in $m^3 (kg.mol)^{-1} s^{-1}$

	K	A	B
$N_2 + O \rightleftharpoons NO + N$	K_1	$1.8e^{11}$	-38370
	K_{-1}	$3.8e^{10}$	-425
$N + O_2 \rightleftharpoons NO + O$	K_2	$1.8e^7$	-4680
	K_{-2}	$3.8e^6$	-20820
$N + OH \rightleftharpoons NO + H$	K_3	$7.1e^{10}$	-450
	K_{-3}	$1.7e^{11}$	-24560

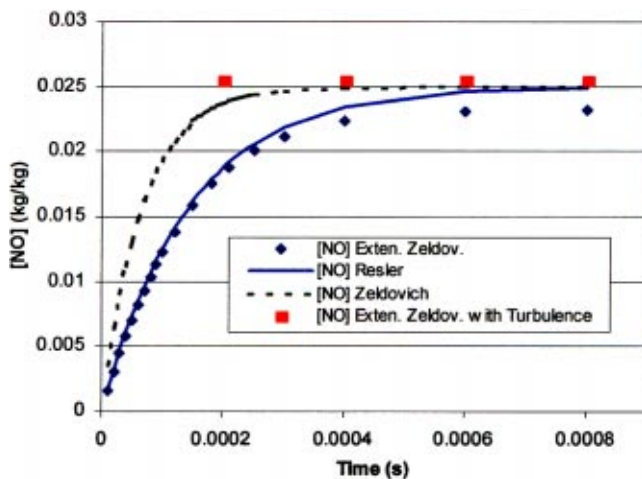


Fig. 5 Equilibrium and relaxation time comparisons of NO models

Table 7 Injected hot gas properties

Hot gas temperature=2278 K	Injection duration= $6.5e^{-5}$ s
Injection velocity=750 m/s	Density= $2.6532 kg/m^3$
Turb. kinetic energy= $1.500e^4 m^2/s^2$	Turb. dissipation= $1.308e^9 m^2/s^3$
Concentration (kg/kg)	
$[C_3H_8]=0.0$	$[O_2]=0.0441$
$[H_2O]=0.0795$	$[N_2]=0.7307$
$[mixture\ fraction]=0.0485619$	$[CO_2]=0.1457$
	$[NO]=0.0$

stated level of closure is reported, [38], to have some known drawbacks, notably in expressing convection by turbulence and turbulence generation.

A single step global reaction is assumed, where CO_2 and H_2O are produced in the background of N_2 . R_{kin} is the reaction rate, which is calculated from chemical kinetics using an Arrhenius-type expression Eq. (2).

$$R_{kin(C_3H_8)} = -AM_{C_3H_8}T^\beta \left[\frac{\rho m_{C_3H_8}}{M_{C_3H_8}} \right]^m \left[\frac{\rho m_{O_2}}{M_{O_2}} \right]^n e^{-E_a/RT} \quad (2)$$

For propane, [39], constants in the rate expression (2) are $A = 4.83e^9 (kmol/m^3)^{1-m-n}$, $\beta = 0.0$, $E_a = 1.26e^8 (kJ/mol)$, $m = 0.1$, $n = 1.65$. These kinetic constants, including the pressure dependency, can also be estimated by calibrating computed and measured laminar speeds over a range of equivalence ratios, pressures, and temperatures, [34,40].

2.6 NO_x Emission Model. NO_x emissions in SI and CI engines, propelled by the current environmental regulations are widely studied in literature. A wave rotor design for a high TIT gas turbine cycle with NO_x constraints has been done, [41], ignoring the NO_x formation and kinetics in the wave rotor component. For the selected deflagrative cycle, only the thermal NO production is taken into account, governed by the three extended Zeldovich steps, Table 6, [42,43]. Partial equilibrium assumptions are made to obtain the concentrations of radicals O, OH, and H, since they are not present in the combustion model.

NO history, obtained with this model, for a given temperature, and initial O₂ and N₂ concentrations, is compared in Fig. 5 with the two-equation Zeldovich [44] and Resler's [45] models. For this transient test case, the relaxation times of the extended Zeldovich mechanism and the later model are found to be matching.

The prompt and turbulent NO_x models, although being available, are postponed to the future work. The contribution of prompt NO becomes significant in rich flames only. Turbulence-chemistry interaction is expressible by the joint PDF of concentrations and temperature. For the transient equilibrium test case, Fig. 5, it is observed that the results are dependent to the number of PDF integration steps. For this reason turbulent contribution will not be included until a complete validation study is available.

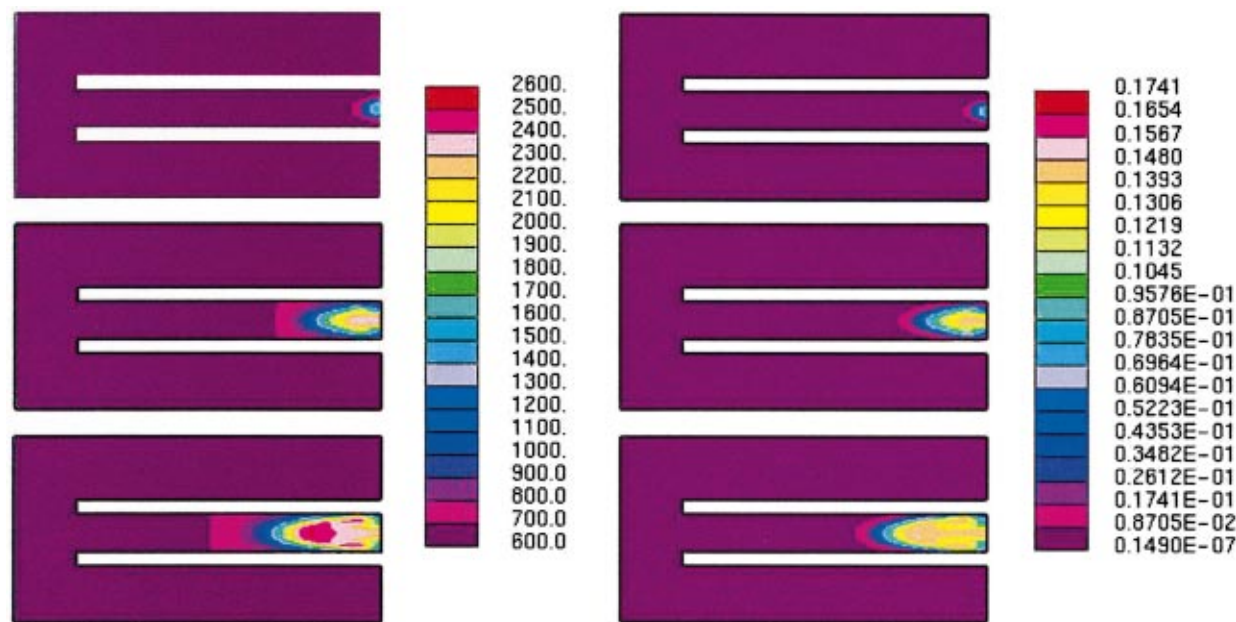


Fig. 6 Temperature (K) (left) and CO₂ concentration (kg/kg) (right) during the hot gas injection phase. Time=5e-6, 4e-5, and 6.5e-5 s.

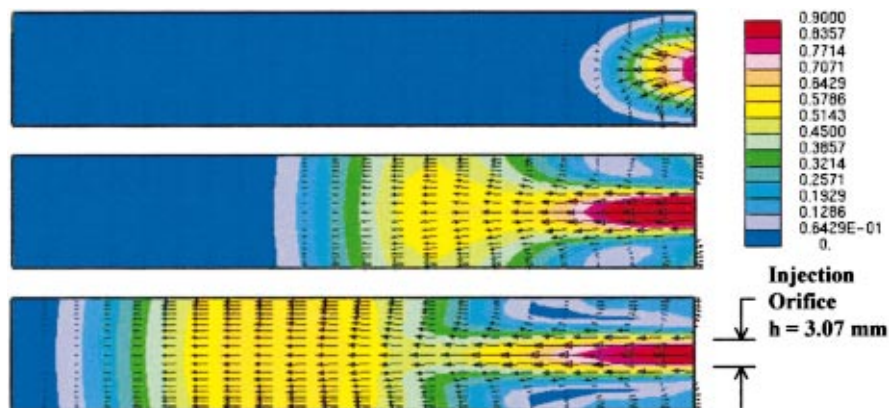


Fig. 7 Flow and Mach number in the fuel partition during the hot gas injection phase. Time=5e-6, 4e-5, and 6.5e-5 s.

3 The Ignition Phase

Two different methods of ignition are investigated. In one case, combustion is initiated in the so-called ignition cells, by specifying percent fuel burned during the ignition process and time it takes, [25]. Although simple to implement, this method resulted in unrealistically long ignition durations. This further decreased the rate of subsequent combustion due to rapidly decaying turbulence level in the channel.

Another approach is to use hot gas from previously combusted products to initiate combustion, which is one of the viable alternatives of the practical experimental wave rotor engines, [11,29]. In this study, hot gas, Table 7, is introduced to the computational domain from a small surface at the inlet, at a specified mass flow rate and chemical composition. Temperature, CO_2 , and Mach number distributions at three time steps during this period are presented in Figs. 6, 7.

With these hot gas injection properties it became possible to obtain a partition exit flow that is close to the critical state, Mach number reaching to 1.0 as shown in Fig. 8. High speed flow at the partition exit is beneficial because of the corresponding penetration and high rate of turbulent energy production at the jet mixing layer, which in parallel increases the reaction rate in the channel, Fig. 9.

At this level, in ICWR engines, detailed ignition models are not attempted. However, future work can utilize fuel spray and special models implemented in existing codes for engine combustion, [40,46,47].

4 Initial Fuel Distribution in the Channel

Configurations with different initial fuel distributions can be achieved by stratification in the inlet duct.



Fig. 8 Mach number distribution of partition exit flow. Hot gas injection. Time=0.165 ms. (Refer to Fig. 7 for color map.)

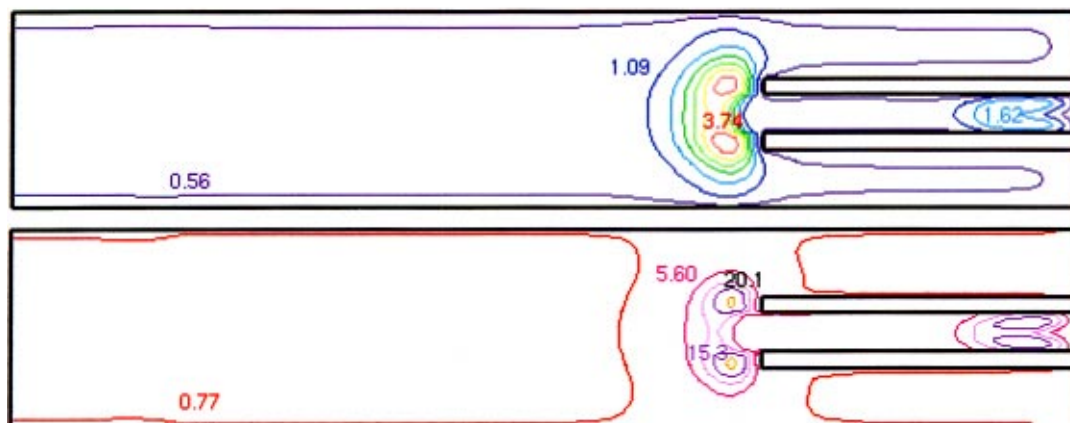


Fig. 9 Nondimensional turbulent kinetic energy (top) and dissipation (bottom) distributions of partition exit flow. Hot gas injection. Time=0.165 ms.

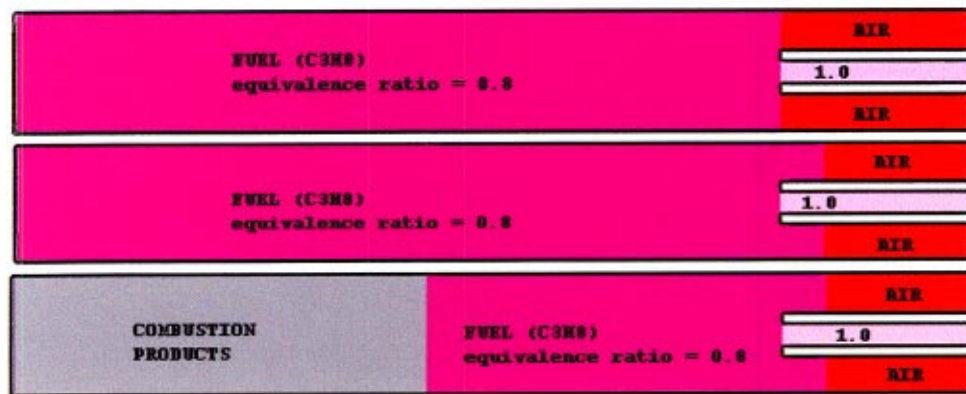


Fig. 10 Initial fuel distribution. Configurations are labeled (a), (b) and (c), from top to bottom.

The first configuration that was studied is shown in Fig. 10(a), where a stoichiometric fuel/air is specified between the pilot partition walls, and air with no fuel is present at the top and bottom partitions. The remaining part of the channel is filled with a lean fuel/air mixture of 0.8 equivalence ratio.

Compositions of the five-component mixture for sample equivalence ratios are presented in Table 8. For Configuration (a), it is found that due to the severity of the pilot partition exit flow, recirculation regions are developed near the tips of the partition walls. This flow structure abruptly quenched the combustion zone by convecting cold air from the buffer layers, Fig. 11. The reaction rate reduced considerably. For cases where pilot partition walls are not present the reaction rate is initially slow but there is no abrupt termination.

Decreasing the extent of the cold air buffer as shown in Fig. 10(b) has solved this problem. It is observed that, a fairly one-dimensional flame front propagates with high speed up to the

middle of the channel. Approximately at the third millisecond, having slight variations in its value depending on the initial turbulence level and hot gas injection parameters, a sharp slowdown occurs in the combustion rate. The mean velocity in the constant-volume channel oscillates between zero and approximately 60 m/s, as the waves continuously reflected from the boundaries. Dissipation produced at the boundary layer is periodically convected away from walls, increasing reaction rate at core regions. If a considerably long time passes, this results in a ladder like shape in the fuel concentration distribution; a typical plot is shown in Fig. 12.

This slowing of the burning process, which can be associated with the rapidly decaying initial turbulence level in the channel, is not desirable in existing practical wave rotor configurations. In cycle designs studies, especially in long channels with turbulent flame speeds, combustion can be started also from the other channel end by hot combustion products, which normally remain from

Table 8 Initial mixture compositions (kg/kg) and stagnation conditions

Stagnation pressure = 7.6×10^5 Pa Lean ($\Phi = 0.8$)			Stagnation temperature = 600 K Stoichiometric ($\Phi = 1.0$)		
$[C_3H_8] = 0.04856$ mix. frac = 0.04856	$[O_2] = 0.22074$ $[N_2] = 0.73068$	$[CO_2] = 0.0$ $[H_2O] = 0.0$	$[C_3H_8] = 0.05997$ mix. frac = 0.05997	$[O_2] = 0.21809$ $[N_2] = 0.72194$	$[CO_2] = 0.0$ $[H_2O] = 0.0$

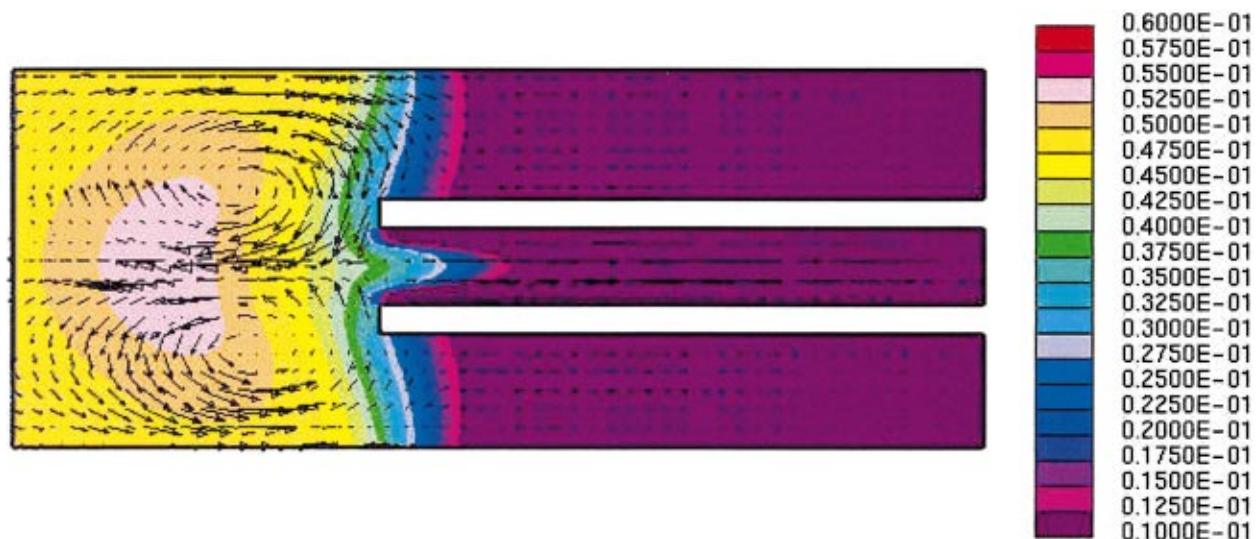


Fig. 11 Pilot fuel partition tip recirculation region. Velocity vectors and C_3H_8 (kg/kg) contours. (Ignition cells.) (Time = 15.5 ms.)



Fig. 12 Fuel (C_3H_8) Concentration in the channel after a long time (10.1 ms). Refer to Fig. 11 for field plot color map.

the previous cycle, [4]. The dual initiation of combustion speeds the process, and initial charge can be consumed in a reasonable time.

Figure 10(c) shows the initial fuel and combustion products distribution for the dual ignition configuration. Compositions of combustion products will be discussed in Section 5.3. Their exact location is determined so that wave rotor channel contains the same amount of fuel as specified in [4] for the design operating conditions of the topped engine at the conditions given in Table 8. This final configuration is the starting point of the selected test cases that are presented in the next section.

5 Results

5.1 Initial Turbulence. In Fig. 13 total channel fuel mass histories are compared for the eight initial turbulence test cases. At the end of the fixed constant volume combustion period, there is approximately 20% variation in the remaining unburned fuel mass, if the extreme case, with unrealistically high turbulent kinetic energy and relatively low dissipation (Case t6), is excluded. As long as a representative trend is used, like the combustion parameters that are kept at their basic literature values, this variation will not affect the conclusions of this study, which focuses on specific design features to improve and quantify NO formation on a relative basis.

The turbulent time scale, which, due to the eddy dissipation model, is in the same order of the turbulent combustion rate, is changing in the course of combustion. Dissipation determines the

later values of turbulent kinetic energy. With a high initial dissipation, initial kinetic energy will die out rapidly, affecting the rate of combustion in later times. During startup, fuel mass consumption is high for the test cases with high initial turbulent kinetic energy. The latter variation is attributed to the flow and to the chemical kinetic time scale term in the reaction rate of the combustion model. It is also observed that resulting combustion front shapes were somewhat different for different initial turbulence levels.

For the NO computation results in this report, the empirical study of Tabaczynski [28] is utilized (Case 1b), since this model appears representative of all the fuel burn rate trends.

5.2 Test Cases. Test cases are selected to analyze NO_x production in the channel for different stratifications and compare the effects of partition, equivalence ratio and air buffer, Fig. 14.

Case (3), has the simplest charge distribution, with no stratification, with lean equivalence ratio. All partitioned cases have the cold air buffer near the hub/shroud inlet, resulting in nearly the same amount of initial fuel. For cases (2a), (2b), and (1g), without having pilot partition walls, it is not realistically possible to introduce such a sharp divided initial air buffer. These cases are studied only to understand separately the effect of fuel stratification and the effect of physical partitioning.

Cases (1f) and (1g) feature air-core region that reduces the remaining unburned fuel in the channel at the end of the constant volume combustion phase. These cases are designed to show the hypothetical limit of emission levels and combustion efficiency

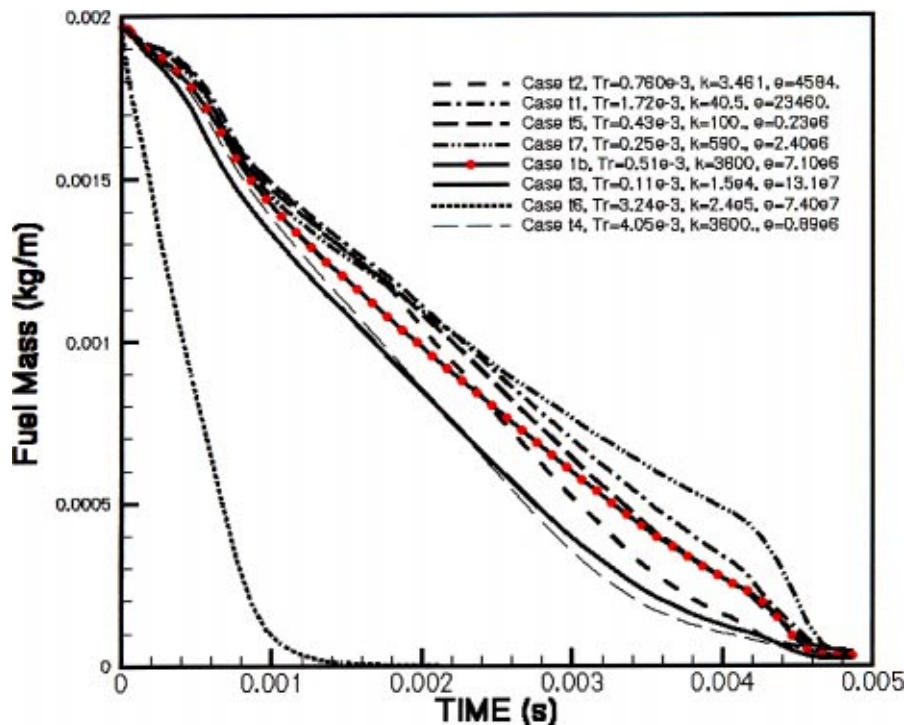


Fig. 13 Total fuel history in the channel for different initial turbulence levels. Dotted curve is specified in the computations.

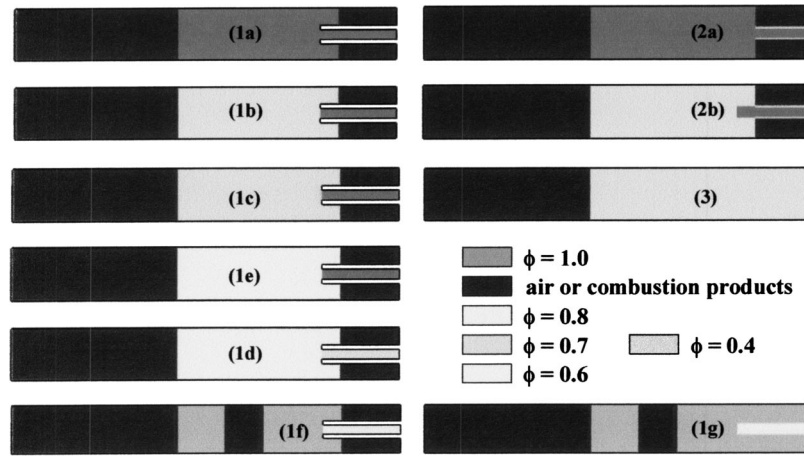


Fig. 14 Test cases studied. Initial fuel distribution. (See color key for equivalence ratios. Cases on left have partition walls).

that can be possible for the configuration and wave rotor cycle studied. Case (1g) does not have pilot partition walls and the air buffer region is not placed for comparison purposes, unlike Cases (2a) and (2b).

5.3 Combustion Products From Previous Cycle. The composition of combustion products that remained from the previous cycle is different for all cases. A reasonable estimate for the initial gas temperature and NO concentration is to use the constant enthalpy/pressure chemical equilibrium composition value and corresponding flame temperature. These parameters cannot be known exactly unless the converged periodic solution of the full cycle is available.

Sensitivity of results to the initial burned gas flame temperature is found to be high, especially for lean mixtures. For stoichiometric test cases with different initial combusted gas temperatures

and zero initial NO content, final NO levels at the end of combustion are similar, but having quite different histories, Fig. 15. Lean cases may approach the same level after very long combustion times.

Compared to the initial equilibrium temperature, initial NO concentration in the channel has a slight effect on the final emission levels. In a global sense, different initial NO levels, with similar combusted gas flame temperatures, will converge as the combustion proceeds.

5.4 Primary Chamber Stoichiometry. Cases 1a, 1b, 1c, and 1e, all have pilot partition walls, with stoichiometric initial composition in the partition. Primary chamber region contains stoichiometric (1a) and lean mixtures with decreasing equivalence ratios (1b-0.8, 1c-0.7, 1e-0.6).

Figure 16 shows the overall NO concentration in the channel

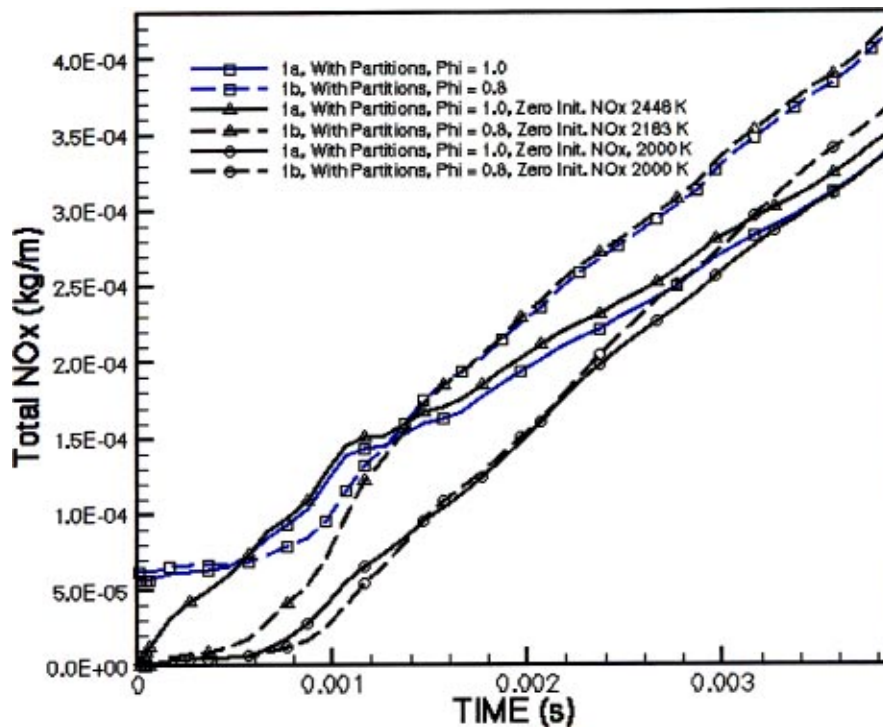


Fig. 15 Total NO mass versus time for different initial combusted gas temperatures and NO levels

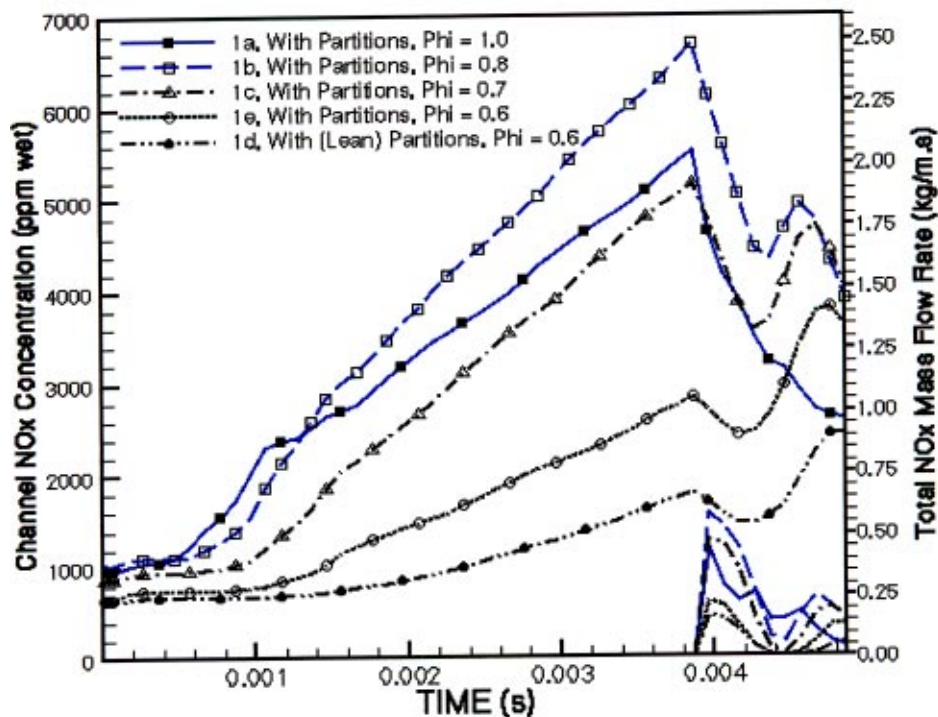


Fig. 16 Total NO_x (ppm) versus time for cases (1a, 1b, 1c, 1e, 1d). (NO_x mass flow through the exhaust port are plotted on lower right corner.)

vs. time for these cases. Wave rotor passage averaged NO concentration reaches its peak around equivalence ratio 0.8 and becomes lower for leaner test cases. This is an expected result in parallel with that of IC engine emission trends. The calculated overall NO_x (ppm) levels are found to be higher than typical limits set by regulations and attained emissions of industrial gas turbines, again as expected.

The turbine exhaust starts at 3.865 ms. During this phase, for some time, NO concentration in the channel decreases. For stoichiometric burn case, this trend continues for the rest of the exhaust phase. But for all lean burn cases, NO concentration in the channel starts increasing again, and for the leanest case, exceeds the level reached at the end of combustion. The initial decrease is attributed to the high NO_x convected by the outflow of previous cycle combustion products. The second peak is due to the relative magnitudes of the three main NO_x sources; ongoing combustion process close to the partition tips, combustion process at the fuel rich mixture/hot products interface and NO_x produced by the combustion products of the previous cycle due to increasing temperature.

For the lean case, the turbine exhaust phase ends and the outlet port is closed before the gas in the channel with high NO concentration reaches to the outlet port. See Fig. 16, where NO mass flow rates through the exhaust port are plotted at the lower right corner. The channel NO concentration begins increasing as the unburned fuel leaves the channel. During this phase of outflow, the channel loses very little NO, since only unburned fuel leaves. Moreover previous combustion products have already left, leaving only the highly expanded partition tip gas, as the sole NO source in the channel. Combustion and complex kinetics at the partition tip region is believed to be the cause, in spite of the temperature decrease due to the expansion wave, which should retard the NO production.

5.5 Effect of Partitions, Cases (1a, 1b, 2a, 2b, 3). Figure 17 shows the channel total fuel mass histories. Fuel consumption rate is higher for stoichiometric, compared to lean mixtures. How-

ever their difference disappears towards the end of constant volume combustion process. In (2a) and (2b), cold air buffer region is in direct contact with the stoichiometric fuel mixture. The fresh air buffer dilutes the combustion zone, and results in very low combustion rates and low NO concentrations, Fig. 18.

The other limiting configuration is the case with no air buffer and partition, Case 3. This case has resulted in the highest emission level of all the configurations studied, expected for a configuration with no possibility of cold air suction to the combustion zone. For Case 3, channel NO concentration peak that occurs during the exhaust phase is also the highest of all the configurations studied, Fig. 18.

Finally, during the exhaust phase, for cases without partition but with air buffer, total NO_x concentration continuously decreases without changing its trend, Fig. 18.

5.6 Pilot Partition Stoichiometry, Cases (1d and 1e). Cases 1d and 1e both have same lean primary chamber composition, with 0.6 equivalence ratio. However in Case 1d, the fuel injection partition region has a leaner composition (with $\Phi = 0.8$) than Case (1e), which is stoichiometric. Referring to Figs. 16 and 17, lean fuel partitions case has slower overall reaction rate with a large quantity of remaining unburned fuel and corresponding low NO concentrations.

5.7 Lowest Emission Design, Cases (1f and 1g). The NO emission for a given engine is application dependent. Parameters like the power requirement, thrust level, specific fuel consumption, type of fuel, turbine inlet temperature limit, etc., influence the lowest available emission limits. A typical new aircraft research goal was the limit of 5 g/kg_{fuel} set as a target of NASA's High Speed Civil Transport program. The emission standards set by the Environmental Protection Agency (EPA) depend on the engine's power output and intended use (utility or industrial). The typical NO_x limit for industrial units with power outputs below 30 MW is 150 ppm (dry), [48]. For some geographic regions, for low

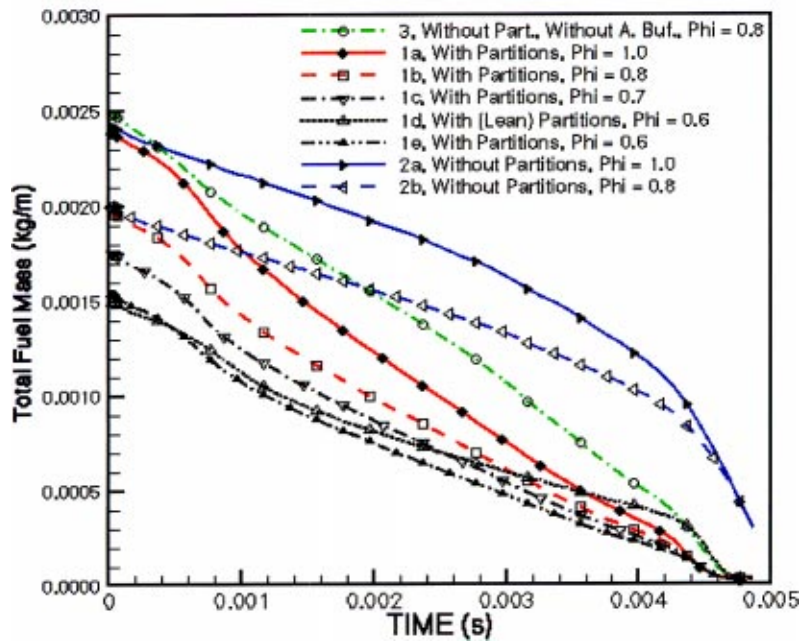


Fig. 17 Unburned fuel mass in the channel

power levels, NO_x limit has been set to 9 ppm (dry at 15% O_2), which is determined by considering the best available technology for gas turbine engines, [49].

For internal combustion wave rotor engines the best available technology is not yet known. For the concept studied in this report, configurations that produce lowest emissions can be designed. Cases (1f) and (1g) are specially selected to show the qualitative NO limits possible. The air core region of these cases reduces the overall fuel-air ratio considerably. Since for these test cases the resulting flame temperature is very low ($T_{\text{flame}}(1f) = 1390 \text{ K}$, $T_{\text{flame}}(1g) = 1479 \text{ K}$), initial equilibrium NO , that is contained only in the previous cycle combustion products, remains practically constant during the combustion.

For these cases the beneficial effect of partition walls to the combustion rate becomes more apparent compared to other cases, since as the average flame temperature drops, combustion rate depends more on turbulence than kinetics. In test Case (1g), which does not have pilot partitions, approximately half of the initial fuel remains unburned in the channel at the start of the turbine exhaust phase. On the other hand, due to the presence of partitions, in test Case (1f) remaining unburned fuel is comparable to other lean cases that are previously discussed.

Case (1f) can be considered as the lowest emission design that will lead to emission targets around 100 ppm, or under $1 \text{ g/kg}_{\text{fuel}}$. The goal of 10 ppm requires more detailed level of understanding.

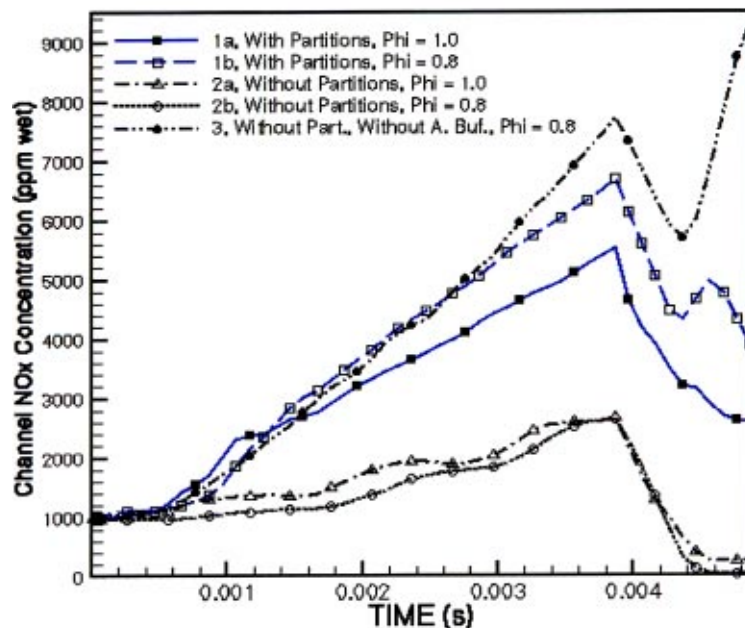


Fig. 18 Total NO_x (ppm) versus time for cases (1a, 1b, 2a, 2b, 3)

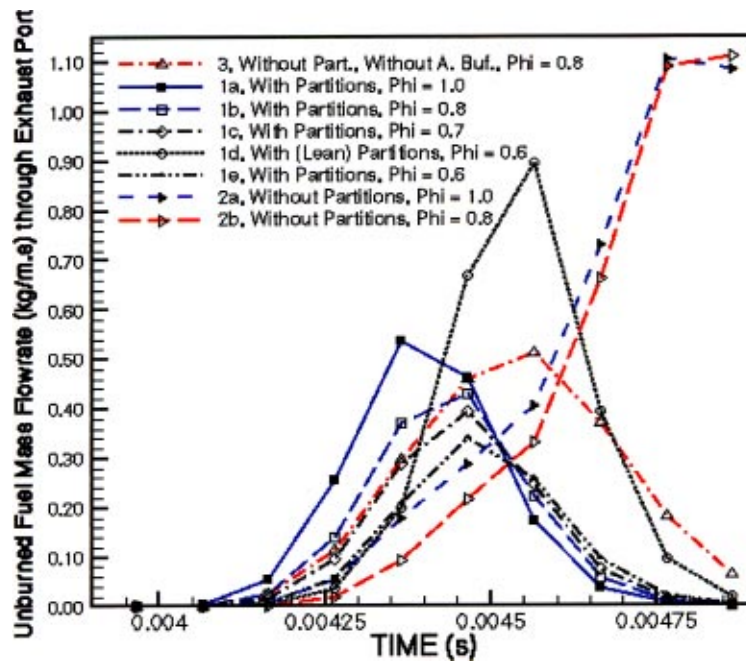


Fig. 19 Unburned fuel exhaust flow rate histories

These levels can be approached with lean premixed combustion but should be noted that additional kinetic mechanisms and interactions with turbulence will occur.

5.8 Turbine Inlet Compositions. The mass flow rate averaged turbine inlet temperatures are not calculated. During exhaust, due to the expansion wave, combustion product's temperature decreases from 2500 K to 1900 K for Case (1c).

Fuel mass flow rate to the turbine during the exhaust phase is plotted in Fig. 19. The fuel spot, which remained from the uncompleted combustion, causes a fuel flux peak as it exhausts from the channel outlet.

The peaks in the instantaneous NO_x concentrations at the exhaust port, in Fig. 20, are due to the channel regions with different temperatures and compositions. Two primary peaks are clearly observed for lean burn configurations corresponding to two combustion zones, surrounding the unburned fuel. These two peaks are not very noticeable for the stoichiometric case. From a design point of view the magnitude of both peaks should be made equal and decreased.

Total ppm of NO exhausted to turbine can be calculated by a mass flow weighted time average. Corresponding values for the test cases are presented in Table 9.

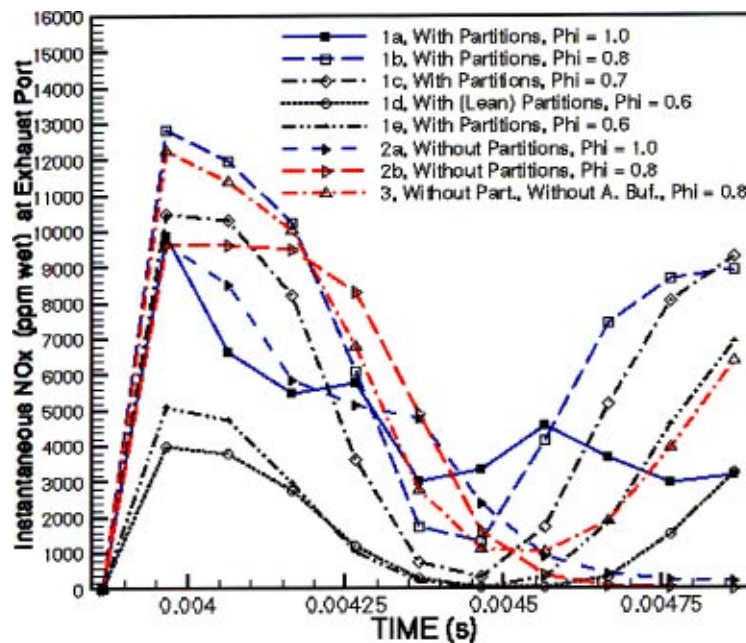


Fig. 20 Instantaneous exhaust NO_x concentrations (ppm wet)

Table 9 Emission indices (EI), mass weighted time-averaged turbine NO_x concentrations (ppm wet), and stagnation pressure rise ratio for combustion

Case	Fuel/Air Ratio	Average Pressure Rise	EI (kg/kg)	NO _x (ppm)	Configuration
1a	0.057	2.7	0.141	5100	(1.0) with partitions
1a_o	0.057	2.7	0.176	5300	zero initial NO, (1.0), with partitions
1b	0.046	2.5	0.216	7100	(0.8) with partitions
1b_o	0.046	2.5	0.257	7200	Zero initial NO, (0.8), with partitions
1c	0.041	2.4	0.185	5400	(0.7) with partitions
1e	0.035	2.2	0.108	2400	(0.6) with partitions
1d	0.035	2.0	0.070	1600	(0.6) with lean (0.8) partitions
1f	0.021	1.6	0.00106	220	Air core, with lean (0.6) partitions
1g	0.023	1.7	0.000378	300	Air core, without air buffer and partitions. With (0.6) fuel region
2a	0.058	2.1	0.094	3700	(1.0) Without partitions
2b	0.045	1.9	0.118	4200	(0.8) Without partitions
3	0.051	2.6	0.229	5700	(0.8) Without air buffer and partitions

6 Conclusions and Future Work

The use of pilot partition to boost the main chamber combustion, and air partitions, in avoiding hot gas leakage to the cavity is apparent. The drastic quenching effect of the air buffer limits the application of this concept, while its amount accurately controlled to cases with pilot partitions only.

Emission indices, calculated at the end of the constant volume combustion phase are tabulated in Table 9 (EI = (kg NO emitted)/(kg fuel burned)). For different configurations, calculated NO emission indices should be compared considering the percent reaction completed. For example “without partition and with air buffer” cases have the least emissions, since the overall chamber temperature had not had time for drastic NO production. Given more time, these cases will also reach or may exceed (as observed in Case 3) the highest emission levels of Table 9.

Since the remaining unburned and initial fuel amounts are different for each configuration, comparison of engine NO_x potential should be based on both the total ppm NO_x exhausted and emission index at the end of constant volume phase, Table 9. Note that Case 3, without partition and air buffer, has the highest emission index of all the configurations (excluding cases with no initial NO in the residual gas). However, because of the absence of partitions that increases the reaction rate, at the end of constant volume combustion, fuel burned is low thus the total ppm NO_x exhausted to turbine is low. The NO formed during a cycle and the NO emitted are not directly linked; a significant amount remains with the residual gas, and affects how much will be formed in the next cycle.

In this research, a particular method for obtaining low NO_x levels in internal combustor wave rotors is studied. As for other combustion devices there may be other design approaches not yet studied for wave machines. In conventional gas turbine combustors the common method for obtaining low emissions is the addition of more air into the primary combustion zone to lower the flame temperature, facilitate better mixing, and eliminate hot spots and reduction of combustor residence time. The stratified combustion approach that is studied here is the appropriate equivalent for non-continuous flow machines.

While comparing the emission levels that are stated in this report with other classes of machinery such as gas turbines and internal combustion engines, the specific application should be taken in to account. Most of the test cases in this study have very high fuel/air ratio compared to the existing applications. This re-

sults very high channel temperatures, high power output, and thus high emissions. Only the “limit” test Cases ((1f) and (1g)) have low fuel/air ratios like conventional engines, Table 9. In parallel, these cases have very low emission indices compared to other cases. The design of the lowest emission internal combustion wave rotor could be realized by a systematic development approach as has been done for conventional gas turbine combustors. A realistic target, perhaps a 10 ppm limit, may be set for wave rotor machines, however only after a 100 ppm engine has been successfully designed and built. For some preliminarily numerical test cases, it is shown in this report that engines close to the 100 ppm goal can be realized.

Recently, multistep reactions for rich combustion allowing fuel-rich pilot partition, [50], and radial leakage flow to the rotor cavity, [51], are studied for the same geometry. Future work may include different NO_x models, unsteady heat transfer, realization of the complete cycle described in Section 1 and comparisons with one-dimensional codes, and modeling rotor swirl in hot gas and fuel injection processes.

Acknowledgments

This work was supported by NASA Glenn Research Center under grant NAG3-2207, and by the Indiana 21st Century Research and Development Fund.

Nomenclature

A	= Arrhenius constant
A	= area
A_{ebu}	= eddy breakup model constant
B_{ebu}	= eddy breakup model constant
β	= temperature exponent
E_a	= activation energy
ε	= dissipation
ϕ	= equivalence ratio
h	= chamber height
η_v	= volumetric efficiency
k	= turbulent kinetic energy
K	= rate constants
l	= turbulent length scale
L	= characteristic length
m	= mass fraction
M	= molecular weight
R	= gas constant
ρ	= density
t	= time
T	= temperature
T_i	= turbulent intensity
T_R	= reaction time scale
T_{tb}	= turbulent time scale
τ_e	= eddy decay time
u'	= fluctuating velocity component
$\bar{U}$	= average velocity
y^+	= nondimensional first cell distance

References

- [1] 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), p. 469.
- [2] Nalim, M. R., 1994, “Wave Cycle Design for Wave Rotor Engines with Limited Nitrogen Oxide Emissions,” Ph.D. thesis, Cornell University, Ithaca, NY.
- [3] Paxson, D. E., 1995, “A Numerical Model for Dynamic Wave Rotor Analysis,” *AIAA Paper No. 95-2800*.
- [4] Nalim, M. R., 1997, “Numerical Study of Stratified Charge Combustion in Wave Rotors,” *AIAA Paper No. 97-3141*.
- [5] Wilson, J., and Paxson, D. E., 1996, “Wave Rotor Optimization for Gas Turbine Engine Topping Cycles,” *J. Propul. Power*, **12**(4), pp. 778–785.
- [6] Welch, G. E., 1996, “Two-Dimensional Computational Model for Wave Rotor Flow Dynamics,” *Paper No. NASA TM-107192*.
- [7] Larosiliere, L. M., 1995, “Wave Rotor Charging Process: Effects of Gradual Opening and Rotation,” *J. Propul. Power*, **11**(1), pp. 178–184.
- [8] Nalim, M. R., 1999, “Assessment of Combustion Modes for Internal Combustion Wave Rotors,” *ASME J. Eng. Gas Turbines Power*, **121**(2), pp. 265–271.

- [9] Nalim, M. R., 2000, "Longitudinally Stratified Combustion in Wave Rotors," *J. Propul. Power*, **16**(3), pp. 1060–1068.
- [10] 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 No. 98-3399.
- [11] Paxson, D. E., 1993, "A Comparison Between Numerically Modeled and Experimentally Measured Loss Mechanisms in Wave Rotors," Paper No. AIAA-93-2522.
- [12] Jones, S. M., and Welch, C., 1996, "Performance Benefits for Wave Rotor Topped Gas Turbine Engines," Paper No. NASA TM-107193.
- [13] Paxson, D. E., 1996, "A Numerical Investigation of the Start-Up Transient in a Wave Rotor," Paper No. NASA TM-107196.
- [14] Paxson, D. E., 1996, "Numerical Simulation of Dynamic Wave Rotor Performance," *J. Propul. Power*, **12**(5), pp. 949–957.
- [15] Kentfield, J. A. C., 1998, "Circumferential Wave Dividers in Wave-Rotors," Paper No. AIAA-98-3397.
- [16] Freitas, C. J., 1995, "Perspective: Selected Benchmarks From Commercial CFD Codes," *ASME J. Fluids Eng.*, **117**, pp. 208–218.
- [17] Patankar, S. V., and Spalding, D. B., 1972, "A Calculation Procedure for Heat Mass and Momentum Transfer in Three Dimensional Parabolic Flows," *Int. J. Heat Mass Transfer*, **15**, pp. 1787–1806.
- [18] Issa, R. I., 1986, "Solution of the Implicitly Discretised Fluid Flow Equations by Operator Splitting," *J. Comput. Phys.*, **62**, pp. 40–65.
- [19] Abraham, J., Bracco, F. V., and Reitz, R. D., 1985, "Comparisons of Computed and Measured Premixed Charge Engine Combustion," *Combust. Flame*, **60**, pp. 309–322.
- [20] Gran, I. R., Ertesvag, I. S., and Magnussen, B. F., 1997, "Influence of Turbulence Modeling on Predictions of Turbulent Combustion," *AIAA J.*, **35**(1), pp. 106–110.
- [21] Abraham, et al., 1988, "Pressure Non-uniformity and Mixing Characteristics in Stratified Charge Rotary Engine Combustion," SAE Technical Paper No. 880624.
- [22] El Tahry, S. H., 1983, " k - ϵ Equation For Compressible Reciprocating Engine Flows," *J. Energy*, **4**, pp. 345–353.
- [23] Launder, B. E., and Spalding, D. B., 1974, "The Numerical Computation of Turbulent Flow," *Comput. Methods Appl. Mech. Eng.*, **3**, p. 269.
- [24] Magnussen, B. F., and Hjertager, B. H., 1976, "On Mathematical Modeling of Turbulent Combustion With Special Emphasis on Soot Formation and Combustion," *16th Symposium on Combustion*, The Combustion Institute, Pittsburgh, PA, pp. 719–729.
- [25] Hjertager, B. H., 1982, "Simulation of Transient Compressible Turbulent Reactive Flows," *Combust. Sci. Technol.*, **27**, pp. 159–170.
- [26] Pope, S. B., 1990, "Computations of Turbulent Combustion: Progress and Challenges," *23rd Symposium on Combustion*, The Combustion Institute, Pittsburgh, PA.
- [27] Versteeg, H. K., and Malalasekera, W., 1995, *An Introduction to Computational Fluid Dynamics—The Finite Volume Method*, Longman, London.
- [28] Tabaczynski, R. J., 1976, "Turbulence and Turbulent Combustion in Spark Ignition Engines," *Prog. Energy Combust. Sci.*, **2**, pp. 143–165.
- [29] Nalim, R. M., and Jules, K., 1998, "Pulse Combustion and Wave Rotors for High-Speed Propulsion Engines," Paper No. AIAA-98-1614.
- [30] Lau, J. H. W., 1995, "Comparison of PDF and Eddy-Dissipation Combustion Models Applied to a Propane Jet Flame," *Combust. Flame*, **102**, pp. 209–215.
- [31] Libby, P. A., and Williams, F. A., 1980, "Turbulent Reacting Flows," *Topics in Applied Physics*, Vol. 44, Springer-Verlag, New York.
- [32] Raju, M. S., 1992, "Heat Transfer and Performance Characteristics of a Dual-Ignition Wankel Engine," SAE Technical Paper No. 920303.
- [33] Catlin, C. A., Fairweather, M., and Ibrahim, S. S., 1995, "Predictions of Turbulent, Premixed Flame Propagation in Explosion Tubes," *Combust. Flame*, **102**, pp. 115–128.
- [34] Kuo, T. W., and Reitz, R. D., 1989, "Computation of Premixed-Charge Combustion in Pancake and Pent-Roof Engines," SAE Technical Paper No. 890670.
- [35] Kong, S., Han, Z., and Reitz, R. D., 1995, "The Development and Application of a Diesel Ignition and Combustion Model for Multidimensional Engine Simulations," SAE Technical Paper No. 950278.
- [36] Spalding, D. B., 1971, "Mixing and Chemical Reaction in Steady Confined Turbulent Flames," *13th Symposium on Combustion*, The Combustion Institute, Pittsburgh, PA.
- [37] Brizuela, E. A., and Bilger, R. W., 1996, "On the Eddy Break-up Coefficient," *Combust. Flame*, **104**, pp. 208–212.
- [38] Hulek, T., and Lindstedt, R. P., 1996, "Computations of Steady-State and Transient Premixed Turbulent Flames Using PDF Methods," *Combust. Flame*, **104**, pp. 481–504.
- [39] Westbrook, C. K., and Dryer, F. L., 1981, "Simplified Reaction Mechanisms for the Oxidation of Hydrocarbon Fuels in Flames," *Combust. Sci. Technol.*, **27**, pp. 31–43.
- [40] Kuo, T., 1992, "Multidimensional Port-and-Cylinder Gas Flow, Fuel Spray, and Combustion Calculations for a Port-Fuel-Injection Engine," SAE Technical Paper No. 920515.
- [41] Nalim, M. R., Moscarì, J. C., and Resler, L., 1993, "Wave Cycle Design for NO_x Limited Wave Rotor Core Engines for High Speed Propulsion," ASME Paper No. 93-GT-426.
- [42] Flowers, W. L., Hanson, R. K., and Kruger, C. H., 1975, "Kinetics of the Reaction of Nitric Oxide with Hydrogen," *15th Symposium on Combustion*, The Combustion Institute, Pittsburgh, PA, p. 823.
- [43] Monat, J. P., Hanson, R. K., and Kruger, C. H., 1979, "Shock Tube Determination of the Rate Coefficient for the Reaction N₂ + O → NO + O," *17th Symposium on Combustion*, The Combustion Institute, Pittsburgh, PA, p. 543.
- [44] Zeldovich, B., and Raizer, P., 1966, *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena*, Vol. 1, Academic Press, San Diego, CA.
- [45] Nalim, M. R., 1985, "NO_x Control in Natural Gas Engines Using Exhaust Gas Recirculation," M. Sc. thesis, Cornell University, Ithaca, NY.
- [46] Hessel, R. P., and Ruthland, C. J., 1995, "Intake Flow Modeling in a Four-Stroke Diesel Using KIVA-3," *J. Propul. Power*, **11**(2), pp. 378–384.
- [47] Kong, S., Han, Z., and Reitz, R. D., 1995, "The Development and Application of a Diesel Ignition and Combustion Model for Multidimensional Engine Simulations," SAE Technical Paper No. 950278.
- [48] Lefebvre, H. A., 1999, *Gas Turbine Combustion*, Taylor and Francis, London.
- [49] Correa, S. M., 1992, "A Review of NO_x Formation Under Gas Turbine Combustion Conditions," *Combust. Sci. Technol.*, **87**, pp. 329–362.
- [50] Pekkan, K., and Nalim, R., 2002, "On Alternative Models for Internal Combustor Wave Rotor Simulation," Spring Technical Meeting of the Central States Section, Knoxville, TN.
- [51] Pekkan, K., and Nalim, R., 2002, "Control of Fuel and Hot-Gas Leakage in a Stratified Internal Combustion Wave Rotor," Paper No. AIAA-2002-4067.

The Two-Flux Model Applied to Radiative Transfer and Forced Convection in the Laminar Boundary Layer on a Flat Plate

F. Malpica

Departamento de Termodinámica y,
Fenómenos de Transferencia,
Universidad Simón Bolívar,
Caracas AP 89000, Venezuela

N. Moreno

A. Tremante

Departamento de Conversión
y Transporte de Energía,
Universidad Simón Bolívar,
Caracas AP 89000, Venezuela

This study presents the analysis of the thermal boundary layer considering combined convection and radiation in an absorbing, emitting, and scattering medium flowing over a flat plate. At high temperatures the presence of thermal radiation alters the temperature distribution in the boundary layer, which in turn affects the heat transfer at the wall. In many industrial applications, such as in the cooling of turbine and compressors blades, radiative heat transfer plays an important role. The treatment of heat transfer by combined convection and radiation in the boundary layer leads to a set of partial differential and integrodifferential equations, which must be solved simultaneously. The exact solutions are seldom possible and the investigators resort to approximate methods. In the present analysis the two-flux model is used to describe the radiative heat flux in the energy equation. This model reduces the equations that govern the problem to a set of coupled partial differential equations. A finite difference scheme, called "method of columns," is used to transform the resulting equations into an ordinary differential equation system which simplifies the solution. Results for the temperature profile and heat fluxes showed close agreement with the thin and thick limits. The method proposed proves to be useful to investigate the effect of the different radiation parameters on the thermal boundary layer, and also to be accurate enough for engineering applications. [DOI: 10.1115/1.1496775]

Introduction

The more frequent use of high-temperature fluids in modern technological processes has made necessary the study of radiation effects in the analysis of the thermal systems. This way is very useful to extend the study of thermal laminar boundary layers, with the finality to include radiation effects within the fluid, between the fluid and the limiting walls, and through reliable and faster procedures.

Viskanta and Grosh [1] considered the case of an optically thick boundary layer using the Rosseland [2] approximation for a gray gas in forced convection over a wedge. Cess [3] made the analysis of an optically thin boundary layer over a flat plate by separating the temperature profile in two regions: the boundary layer adjacent to the wall, that is optically thin, and a semi-infinite radiation layer, in which the effect of conduction is neglected and the velocity is considered uniform. Oliver and McFadden [4] solved the case of a laminar boundary layer in forced convection over flat plates through the exact formulation that includes the solution of the integro-differential energy equation using an iterative type of solution. Lee et al. [5] considered the case of the thermal boundary layer over an isothermal flat plate by a multilayer approach.

The primary purposes of this study is to present an analysis of the combined radiation and forced convection over a isothermal flat plate of a gray participating medium using the two-flux model. This model is based on the description of the radiation intensity in the medium by only two uniform values (Sidall [6]), one in the forward direction and the other in the backward direction. This simplification converts the equation of radiative transfer into two coupled nonlinear differential equations of second orders (Tremante and Malpica [7]), instead of a nonlinear integro-differential equation (Siegel and Howell [8]). Then the resulting system of

partial differential equations is transformed into a set of two ordinary differential equations using a finite difference scheme (method of columns). The resulting ordinary two-point boundary value problem is solved using a variable order, variable step-size finite difference method with deferred corrections (Pereira [9]). Effects of scattering, optical thickness, the conduction-radiation parameter on the temperature profiles and heat transfer fluxes across the boundary layer are investigated. The encouraging results obtained in the present analysis are compared to those solutions obtained through more complex approaches.

Problem Formulation

This analysis considers the interaction of radiation with forced convection in the laminar boundary layer flow of an absorbing, emitting, isotropically scattering, and compressible fluid over a flat plate. It is assumed that the fluid is a perfect gas and is gray, the viscosity varies linearly with the temperature. The specific heat and the Prandtl number are constant, and the external flow temperature T_∞ is uniform. The surface of the wall is opaque, gray, and a diffuse emitter and reflector. A sketch of the geometry of the physical model and the coordinate system is shown in Fig. 1.

The mathematical model for the assumed physical problem is prescribed by the conservation equation of mass, momentum, and energy. These equations are:

$$\frac{\partial(\rho u)}{\partial x} + \frac{\partial(\rho v)}{\partial y} = 0 \quad (1)$$

momentum

$$\rho \left(u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} \right) = \frac{\partial}{\partial y} \left(\mu \frac{\partial u}{\partial y} \right) \quad (2)$$

energy

Contributed by the Heat Transfer Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the HT Division August 2000; final revision received by the ASME Headquarters April 2001. Associate Editor: H. D. Nelson.

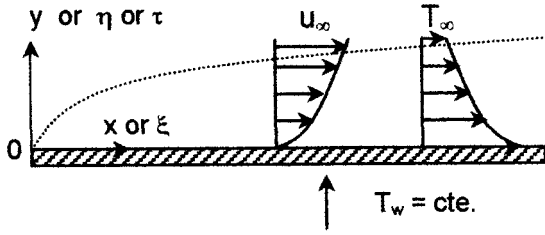


Fig. 1 Geometry of the physical model and the coordinate system

$$\rho C_p \left(u \frac{\partial T}{\partial x} + v \frac{\partial T}{\partial y} \right) = \mu \left(\frac{\partial u}{\partial y} \right)^2 + \frac{\partial}{\partial y} \left(k \frac{\partial T}{\partial y} \right) - \frac{\partial q_r}{\partial y}. \quad (3)$$

The last term in the right-hand side is the divergence of the radiative heat flux, and according to Özisic [10] is given by

$$\begin{aligned} \frac{\partial q_{r,y}}{\partial y} = & \kappa_a \left[(4\sigma T^4 - 2\varepsilon_w E_2(\tau) \sigma T_w^4) \right. \\ & - 4(1 - \varepsilon_w) E_2(\tau) \int_0^\infty \sigma T'^4 E_2(\tau') d\tau' \\ & \left. - 2 \int_0^\infty \sigma T'^4 E_1(|\tau - \tau'|) d\tau' \right]. \end{aligned} \quad (4)$$

It means that the energy Eq. (3) is a nonlinear integro-partial differential equation.

The boundary conditions are

$$u = v = 0, \quad T = T_w \quad \text{at } y = 0 \quad (5a)$$

$$u = u_\infty, \quad T = T_\infty \quad \text{at } y \rightarrow \infty \quad (5b)$$

$$T = T_o(y) \quad \text{at } x = 0. \quad (5c)$$

The foregoing continuity, momentum, and energy equations can be transformed by the application of the standard transformation used in the analysis of the nonradiating boundary layer heat transfer.

We define a stream function as

$$\psi(x, y) \equiv f(\eta) (\nu_o x u_\infty)^{1/2} \quad (6)$$

and a new independent variable $\eta(x, y)$ is defined as

$$\eta \equiv \left(\frac{u_\infty}{\nu_o x} \right)^{1/2} \int_0^y \frac{\rho}{\rho_o} dy'. \quad (7)$$

and a dimensionless temperature as

$$\theta = \frac{T}{T_o}. \quad (8)$$

Then the continuity Eq. (1) is identically satisfied. The momentum and energy Eqs. (2) and (3), are, respectively, transformed to:

$$\frac{d^3 f}{d\eta^3} + \frac{1}{2} f \frac{d^2 f}{d\eta^2} = 0 \quad (9)$$

energy

$$\frac{1}{P_r} \frac{\partial^2 \theta}{\partial \eta^2} + \frac{1}{2} f \frac{\partial \theta}{\partial \eta} = \frac{\partial f}{\partial \eta} \xi \frac{\partial \theta}{\partial \xi} + \xi \frac{\partial Q^r}{\partial \tau} - E_\infty \left(\frac{\partial^2 f}{\partial \eta^2} \right)^2 \quad (10)$$

where various nondimensional quantities are defined as follows:

$$\xi = \frac{4\sigma T_o^3 \kappa_o x}{\rho_o C_p u_\infty} = \frac{(\kappa_o x)^2}{NP_r Re_x} \quad \text{independent variable along the wall} \quad (11)$$

$$N = \frac{\kappa_o k_o}{4\sigma T_o^3} \quad \text{conduction-radiation parameter} \quad (12)$$

$$Re_x = \frac{u_\infty x}{\nu_o} \quad \text{Reynolds number} \quad (13)$$

$$\tau = (\xi NP_r)^{1/2} \eta \quad \text{optical variable} \quad (14)$$

$$Q^r = \frac{q_r}{4\sigma T_o^4} \quad \text{dimensionless radiative heat flux} \quad (15)$$

$$E_\infty = \frac{u_\infty^2}{C_p T_o} \quad \text{Eckert number ([11])}. \quad (16)$$

The radiation term in Eq. (10) is solved using the two-flux model (Sidall [6]). This model is based on the assumption that radiation in the medium may be represented by two uniform fluxes, one in the forward hemisphere and the other in the backward hemisphere (Tremante and Malpica [7]). In principle this formulation replaces the equation of radiative transfer by two-coupled nonlinear partial differential equations, and the expression for the divergence of the radiative flux is given as

$$\frac{\partial Q^r}{\partial \tau} = (1 - w_o)(\theta^4 - G^*) \quad (17)$$

$$G^* = \frac{G}{4\sigma T_o^4} \quad (18)$$

where G^* is the dimensionless total incident irradiation. Finally, the energy equation is reduced to

$$\frac{1}{P_r} \frac{\partial^2 \theta}{\partial \eta^2} + \frac{1}{2} f \frac{\partial \theta}{\partial \eta} = \frac{\partial f}{\partial \eta} \xi \frac{\partial \theta}{\partial \xi} + (1 - W_o) \xi (\theta^4 - G^*) - E_\infty \left(\frac{\partial^2 f}{\partial \eta^2} \right)^2 \quad (19)$$

according to the two-flux model is coupled with

$$\frac{\partial^2 G^*}{\partial \eta^2} = -A \xi NP_r (1 - W_o) (\theta^4 - G^*). \quad (20)$$

The boundary conditions for the momentum Eq. (15) are

$$\eta = 0 \quad f = f' = 0 \quad (21a)$$

$$\eta \rightarrow \infty \quad f' = 1. \quad (21b)$$

The nondimensional boundary condition for the energy Eq. (22) is

$$\eta = 0 \quad \theta = 1 \quad (22a)$$

$$\eta \rightarrow \infty \quad \theta = \frac{T_\infty}{T_w} = \theta_\infty \quad (22b)$$

and for the irradiation Eq. (20) is

$$\eta = 0 \quad \left(\frac{\partial G^*}{\partial \eta} \right)_w = -A (\xi NP_r)^{1/2} \frac{\varepsilon_w}{2(2 - \varepsilon_w)} (1 - G_w^*) \quad (23a)$$

$$\eta \rightarrow \infty \quad \left(\frac{\partial G^*}{\partial \eta} \right)_\infty = -\frac{A}{2} (\xi NP_r)^{1/2} (G_\infty^* - \theta_\infty^4) \quad (23b)$$

with the initial condition

$$\theta = \theta_o(\eta) \quad \text{at } \xi = 0 \quad (24)$$

where $\theta_o(\eta)$ is the solution of the nonradiative case.

The two-flux method is particularized for the model of Milne-Eddington, in that case $A = 3$ (Malpica et al. [12], Tremante and Malpica [7], and Siegel and Spuckler [13]).

The momentum Eq. (15) is uncoupled, and therefore is previously solved, and the functions $f(\eta)$, $f'(\eta)$, and $f''(\eta)$ appearing in the energy equation are determined. The energy Eq. (19) and the irradiation function (20) are solved coupled along discrete fluid lines (ξ) perpendicular to the plate. The intervals of the lines

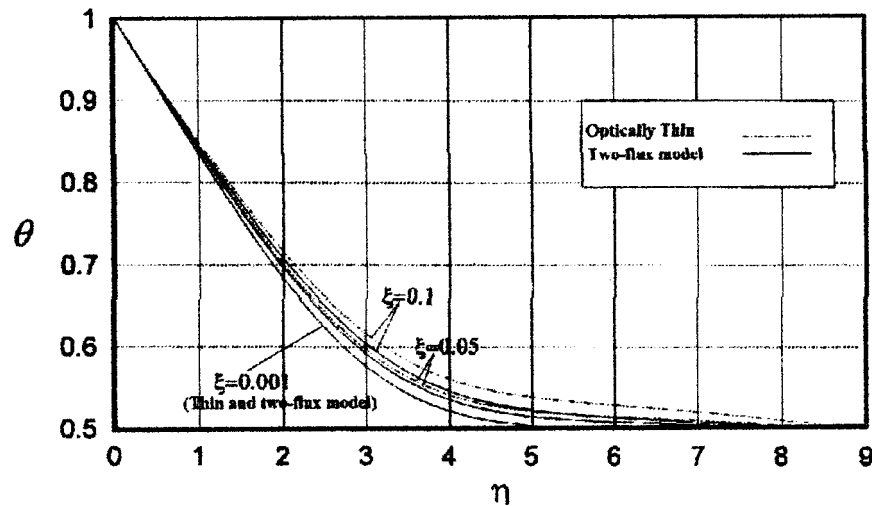


Fig. 2 Temperature profiles over flat-plate forced convection comparison with optically thin approximation. $A=3$, $\theta_\infty=0.5$, $W_o=0.0$, $P_r=1.0$, $N=0.1$, $\varepsilon_w=1.0$.

in the ξ -direction were chosen on a logarithmic basis to better reproduce the initial development of the boundary layer. Upwind values were used in the convective terms according to Patankar ([14]) recommendations. At $\xi=0$, the solution taken corresponds to the nonradiation problem. The algorithm used to solve the set of differential equation is the DVCPR, based on the algorithm developed by Pereira [9], included in the scientific library IMSL.

The asymptotic condition at $\eta \rightarrow \infty$, was verified by checking an error defined as

$$\delta_i^2 = (f_i'')^2 + (\theta_i')^2 \quad (25)$$

satisfying the condition where

$$\delta_i^2 \leq 10^{-6}. \quad (26)$$

Results and Discussion

Before discussing results it is convenient to make some general considerations. First, it is useful to have an expression for the optical thickness of the thermal boundary layer. This expression is obtained by particularizing the value of η in Eq. (14) at the edge of the boundary layer. For the laminar flow without radiation effects and $Pr=1.0$, $\eta_\delta=5$. In general, the optical thickness of the thermal boundary layer is given by

$$\tau_\delta = \eta_\delta \sqrt{P_r N \xi}. \quad (27)$$

For the cases under study the parameter $\sqrt{N\xi}$ characterizes the optical thickness of the thermal boundary layer.

The boundary layer is said to be optically thick when $\sqrt{N\xi} \gg 1$ and optically thin when $\sqrt{N\xi} \ll 1$. Hence at the beginning of the boundary layer where $\xi \ll 1$, the boundary layer tends to behave as a thin gas, and for the asymptotic developed boundary layer, with $\xi \gg 1$, the boundary layer tends to the optically thick limit. The dimensionless parameter N , defined in Eq. (12), determines the relative role of energy transfer by conduction to that by radiation; when $N \gg 1$ radiation is negligible, and when $N \ll 1$ radiation predominates.

The albedo W_o in Eq. (20) is the ratio of the scattering coefficient to the extinction coefficient. When $W_o=0.0$, scattering is not present in the medium and radiation interacts with fluid particles through absorption. Hence the radiation effects are important. When $W_o=1.0$, all the radiation energy is scattered and absorption has no influence in the fluid particles; hence no interaction can be perceived. These previous considerations can facilitate the understanding of the following results.

The solutions are considered for the hot wall condition: $\theta_\infty=0.5$ and $\varepsilon_w=1.0$, $P_r=1.0$, and $E_\infty=0.0$. The parameters N and W_o are varied according to required comparisons.

Figure 2 compares the temperature profiles calculated by the optically thin approximation and those obtained by the two-flux model for a hot wall condition with predominant radiation ($N=0.1$, $W_o=0.0$). It can be observed that at very low values of ξ , existing at the beginning of the boundary layer, both models yield the same results. For the problem considered, the boundary layer may be assumed to be optically thin for $\xi \leq 0.001$. However, as the boundary layer develops and ξ increases, the thin gas model predicts higher temperature values than those given by the two-flux model. This behavior is expected as the thin gas model neglects the attenuation of the energy emitted from the hot fluid particles near the wall. This also explains the larger thickness of the thermal boundary layer for the thin gas approximation. The two-flux model takes into consideration the increasing optical thickness of the boundary layer; therefore its results tend to be more realistic.

Figure 3 shows that for large values of ξ , the two-flux temperature profiles approach those corresponding to the thick gas limit calculated according to reference [1]. For $N=10.0$ no radiation effects are observed. For $N=1.0$ and $\xi=100.0$, the two-flux model temperature distribution presents excellent agreement with the values corresponding to the thick gas model. This means that for $\xi \gg 100.0$ the boundary layer tends to behave as an optically thick gas for the conditions considered. For $N=0.1$, it can be observed that as ξ increases the temperature profiles tend to approximate the optically thick limit. For the largest value of ξ indicated, $\xi=100.0$, the two-flux profile presents only small deviations from the optical thick limit, which means that ξ must be increased even more to be closer to that limit.

To illustrate the effect of scattering, Fig. 4 presents a set of temperature profiles for different values of the albedo W_o . For $W_o=1.0$, pure scattering condition, the temperature distributions coincide with the nonradiating solution for all values of ξ . For $W_o=0.5$, absorption is as important as scattering and radiation becomes significant, therefore, the temperature profiles for each value of ξ tend to separate. For $W_o=0.0$, nonscattering condition, radiation effects are maximum and temperature profiles present the largest separation for the different values of ξ indicated. Hence, as W_o increases radiation effects diminish, and the temperature profiles are shifted toward the nonradiating solution.

Figures 5(a), 5(b), and 5(c) show the effect of radiation on the temperature gradients across the boundary layer calculated by the two-flux model. The results are presented for three values of N :

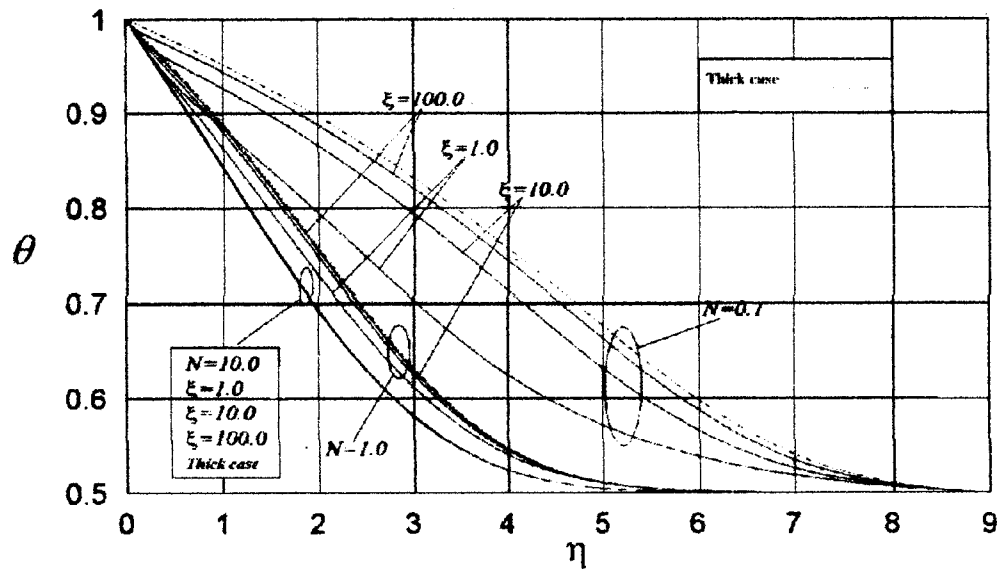


Fig. 3 Temperature profiles over flat-plate forced convection, two-flux method ($A=3$) compared with thick limit. $\theta_\infty=0.5$, $W_o=0.0$, $P_r=1.0$, $N=0.1$, $\varepsilon_w=1.0$.

10.0, 1.0, 0.1. The general tendency of the radiation effect is to reduce the temperature gradients in the region close to the wall as ξ is augmented. This trend is amplified as N decreases. For $N=0.1$ the temperature gradients increase sharply near the wall, as indicated in Fig. 5(c). At the same time radiation tends to smooth the temperature gradients as the fluids separate from the wall for large values of ξ . Viskanta and Grosh [1] also reported this behavior for the thick gas limit. However, the thick gas model does not predict the abrupt reduction of the temperature gradients in the vicinity of the surface. This can lead to serious errors in the calculation of the heat transfer conduction at the wall. Therefore the two-flux model offers a more reliable treatment of the temperature gradients at the wall.

Figures 6(a), 6(b), and 6(c) present the variation of the radiation heat flux across the boundary layer for three different values of N . In these figures it can be observed that as ξ augments the radiation heat flux Q^r -diminishes due to the increase of the optical thick-

ness of the boundary layer. At $\xi=0.001$ as N decreases, Q^r -at the wall tends to approximate to the value $\varepsilon_w/4(1-\theta_\infty^4)$ corresponding to a transparent boundary layer.

The radiant heat flux profiles reaches a maximum at some value $\eta_{\max}$ near the wall and then decreases. In the vicinity of the wall $\partial Q^r/\partial \eta > 0$, which means that the fluid particles are net emitters in this region. Thus the influence of radiation is to impose an effective radiation source within the gas near the surface. Hence, heat transfer by conduction must compensate this net heat transfer by radiation.

Figure 7 shows the total Nusselt Number $Nu_x/R_{ex}^{1/2}$, as a function of ξ for different values of N . All the curves tend to approach the nonradiation case ($Nu_x/R_{ex}^{1/2}=0.332$) as ξ diminishes. It means that the radiation effects on the Nusselt number are not important at the beginning of the thermal boundary layer. As the thermal boundary layer develops and radiation becomes signifi-

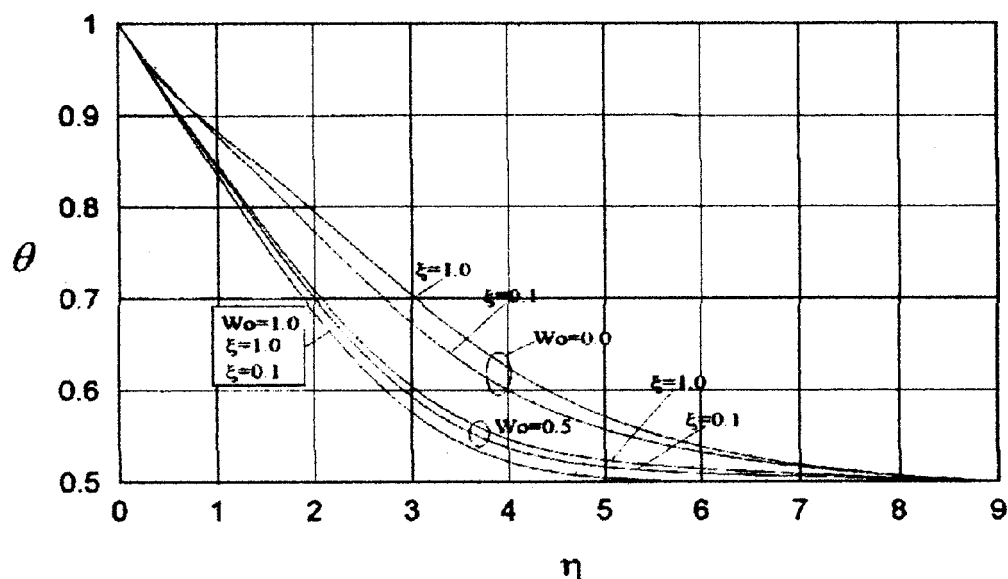
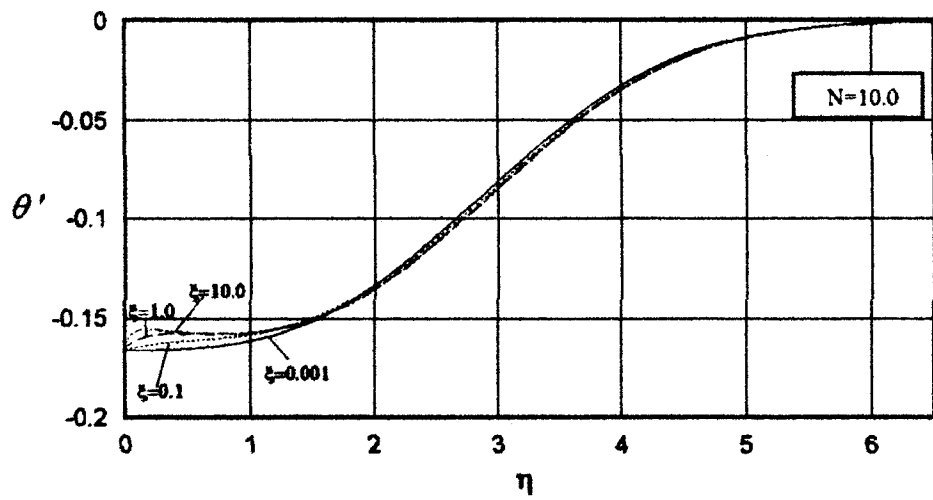
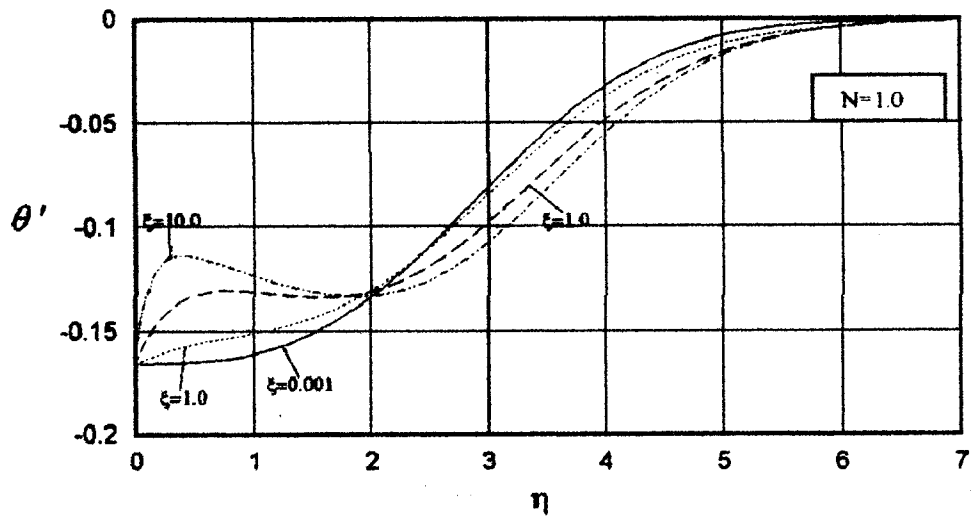


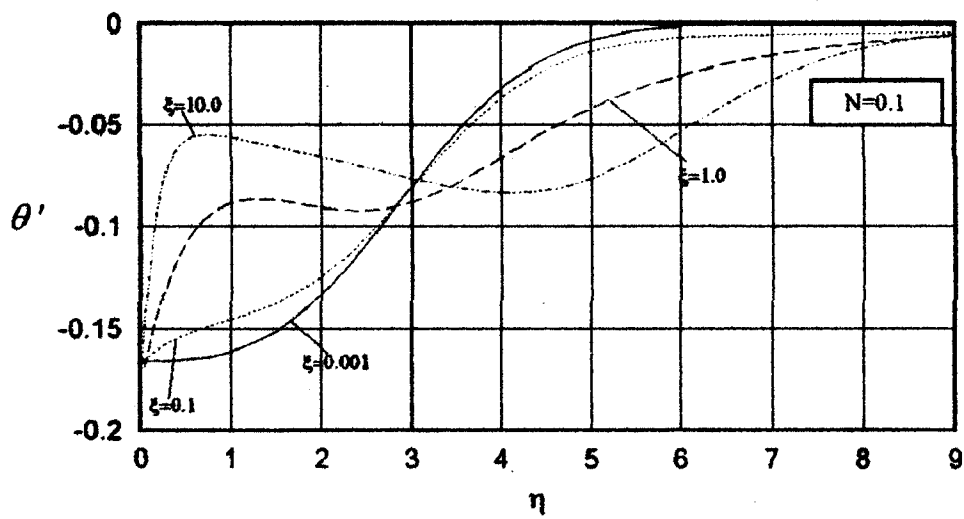
Fig. 4 Temperature profiles over flat-plate forced convection. Effect of scattering, W_o , $\theta_\infty=0.5$, $P_r=1.0$, $N=0.1$, $\varepsilon_w=1.0$, $A=3$.



(a)



(b)



(c)

Fig. 5 (a) Temperature gradients across the boundary layer. Effect of ξ . $P_r=1.0$, $\epsilon_w=1.0$, $\theta_\infty=0.5$, $E_\infty=0.0$, $W_o=0.0$, $A=3$. (b) Temperature gradients across the boundary layer. Effect of parameter ξ . $P_r=1.0$, $\epsilon_w=1.0$, $\theta_\infty=0.5$, $E_\infty=0.0$, $W_o=0.0$, $A=3$. (c) Temperature gradients across the boundary layer. Effect of ξ . $P_r=1.0$, $\epsilon_w=1.0$, $\theta_\infty=0.5$, $E_\infty=0.0$, $W_o=0.0$, $A=3$.

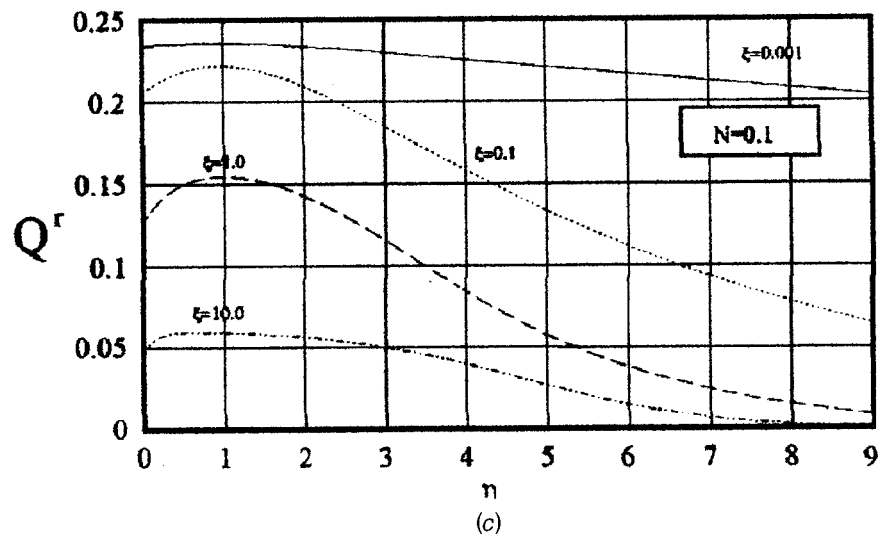
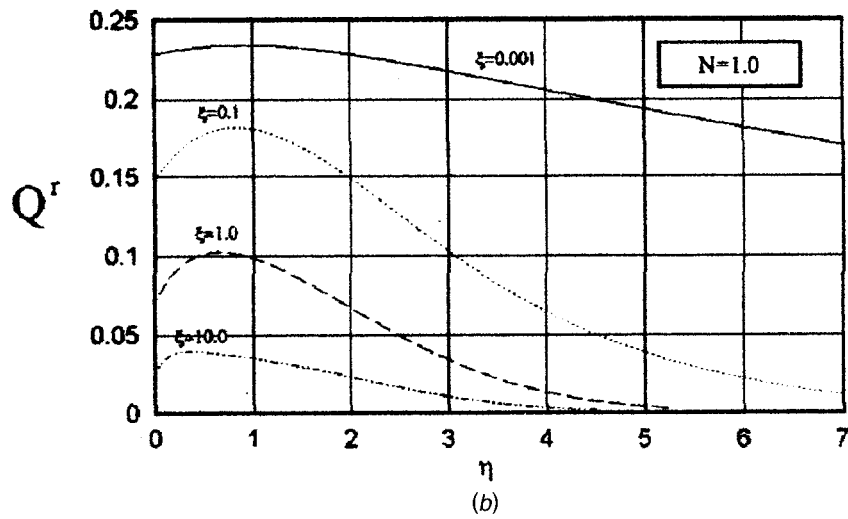
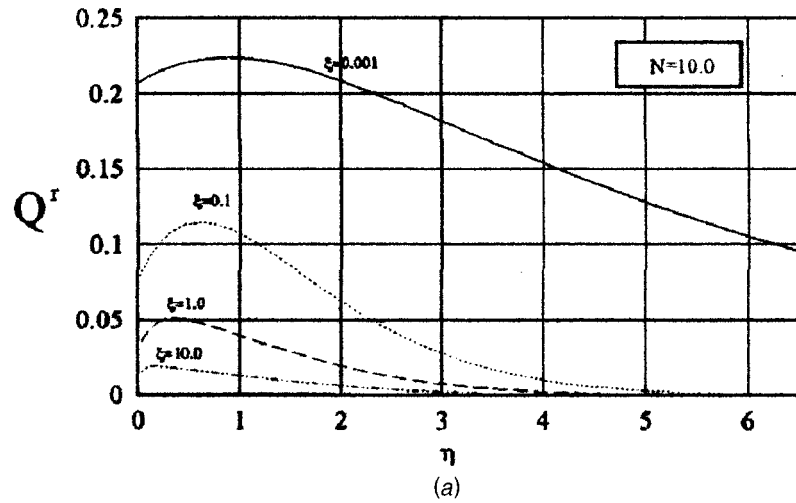


Fig. 6 (a) Effect of parameter ξ on radiative heat flux across the boundary layer. $P_r=1.0$, $\varepsilon_w=1.0$, $\theta_\infty=0.5$, $E_\infty=0.0$, $W_o=0.0$, $A=3$. (b) Effect of parameter ξ on radiative heat flux across the boundary layer. $P_r=1.0$, $\varepsilon_w=1.0$, $\theta_\infty=0.5$, $E_\infty=0.0$, $W_o=0.0$, $A=3$. (c) Effect of parameter ξ on radiative heat flux across the boundary layer. $P_r=1.0$, $\varepsilon_w=1.0$, $\theta_\infty=0.5$, $E_\infty=0.0$, $W_o=0.0$, $A=3$.

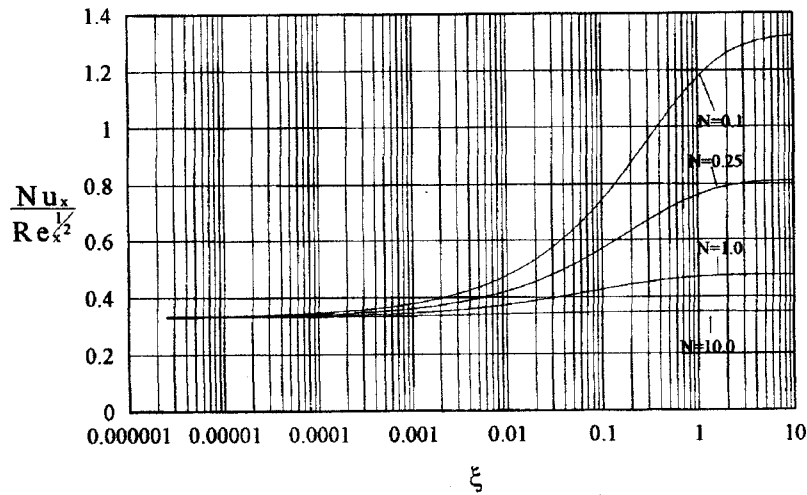


Fig. 7 Total heat flux coefficient (convection+radiation) $P_r=1.0$, $\varepsilon_w=1.0$, $E_\infty=0.0$, $W_o=0.0$, $A=3$

cant, ξ increases and N diminishes, the curves tend to separate from the nonradiation case. It is seen from the results that the effect of radiation is to increase the total heat transfer rate. All the curves present an asymptotic behavior as ξ tends to ∞ (Rosseland limit). Therefore the total heat transfer can be calculated considering radiation and forced convection over isothermal surfaces.

In Table 1, the results for the thermal boundary layer thickness, $(\delta/x)R_{ex}^{1/2}$, are presented for different values of N and ξ . For $N=10.0$, conduction predominates over radiation, and the value tends to 4.9, which represents the nonradiation case. When N is reduced and ξ is increased the thermal boundary layer thickness augments. Its mean that the wall penetration depth into the fluid becomes more important as radiation predominates.

Conclusions

This work shows that the two-flux method for $A=3$ presents an excellent agreement with the limiting cases, optically thin and thick cases, and also allows to calculate temperature distribution and heat transfer across a laminar boundary layer for intermediate values of the optical thickness.

The order of magnitude of the resulting boundary layer thicknesses compare to the nonradiation case, therefore it is possible to conclude that the boundary layer assumptions are valid when radiation effects are significant.

For the hot wall condition the parameter θ_∞ has a negligible influence over temperature and velocity profiles and heat transfer fluxes, because changes in θ_∞ only affect regions far from the surface.

For engineering purposes, as in the case of the interaction between exhaust hot gases and turbine blades, a thermal analysis can be conducted, using the proposed procedure, to predict temperature distribution and heat transfer requirements.

Table 1 Thermal boundary layer thickness $(\delta/x)R_{ex}^{1/2}$

ξ	$N=10.0$	$N=1.0$	$N=0.25$	$N=0.1$
0.00001	4.9	4.9	4.9	4.9
0.0001	4.9	4.9	4.9	4.9
0.001	4.9	4.9	4.9	4.9
0.01	4.9	5.0	5.1	5.1
0.02	4.9	5.1	5.3	5.3
0.05	4.9	5.3	5.7	5.9
0.1	4.9	5.4	6.0	6.3
1.0	4.9	5.4	6.3	6.7
10	4.9	5.4	6.4	6.9

($\varepsilon_w=1.0$, $A=3$, $W_o=0.0$, $P_r=1.0$, $E_\infty=0.0$)

Nomenclature

- A = constant in the two-flux model
- C_p = specific heat a constant pressure
- E_1, E_2 = integro-exponential functions of first and second class, respectively
- E_∞ = Eckert number, based in u_∞
- f = dimensionless stream function
- g = gravity
- G = irradiation
- G^* = dimensionless irradiation
- I = intensity of radiation
- k = thermal conductivity
- N = dimensionless conduction-radiation parameter
- Nu = Nusselt number
- P_r = Prandtl number
- q = heat-flux rate
- R_e = Reynolds Number
- Q = dimensionless heat-flux rate
- T = absolute temperature
- u, v = velocity components in x and y -directions, respectively
- W_o = albedo, scattering parameter
- x = distance along surface
- y = distance perpendicular to surface
- β = thermal expansion coefficient
- δ = boundary layer thickness
- ε = emissivity
- η = dimensionless coordinate in the y -direction
- θ = dimensionless temperature
- ρ = density
- σ = Stefan-Boltzmann constant
- ξ = dimensionless coordinate in the x -direction
- κ = extinction coefficient
- κ_a = absorption coefficient
- κ_b = dispersion coefficient
- τ = optical thickness
- ψ = stream function
- μ = dynamic viscosity
- ν = kinematics viscosity

Subscripts

- c = convection
- k = conduction
- x, y = refers to x and y -direction

o = refers to T_o , reference temperature
 ∞ = refers to free stream
 w = refers to wall

Superscripts

r = radiation
 $+$ = y -direction > 0
 $-$ = y -direction < 0

References

- [1] Viskanta, R., and Grosh, R. J., 1962, "Boundary Layer in Thermal Radiation and Emitting Media," *Int. J. Heat Mass Transf.*, **5**, p. 765.
- [2] Rosseland, S., 1936, *Theoretical Astrophysics*, Oxford University Press, London.
- [3] Cess, R. D., 1964, "Radiation Effects upon Boundary Layer Flow of an Absorbing Gas," *ASME J. Heat Transfer*, **86C**, pp. 469–475.
- [4] Oliver, C. C., and McFadden, P. W., 1966, "The Interaction of Radiation and Convection in the Laminar Boundary Layer," *ASME J. Heat Transfer*, **88C**, pp. 205–213.
- [5] Lee, H. S., Menart, J. A., and Fakery, A., 1990, "Multilayer Radiation Solution for Boundary Layer Flow of Gray Gases," *J. Thermophys. Heat Transfer*, **4**, p. 180.
- [6] Sidall, R., 1972, "Flux Methods for the Analysis of Radiant Heat Transfer," *Proceedings of the Fourth Symposium on Flames and Industry*, The Institute of Fuel, Paper No. 16, pp. 169–170.
- [7] Tremante, A., and Malpica, F., 1994, "Contribution of Thermal Radiation to the Temperature Profile of Ceramics Composite Materials," *ASME J. Eng. Gas Turbines Power*, **116**, pp. 583–586.
- [8] Siegel, R., and Howell, J. R., 1981, *Thermal Radiation Heat Transfer*, 2nd Ed, Hemisphere, Washington DC, p. 501.
- [9] Pereira, V., 1982, "PASVA3: An Adaptive Finite Differences Fortran Program for First Order Nonlinear Ordinary Differential Equation Problems," *Codes for Boundary Problems in Ordinary Differential Equations*, (Lecture Notes in Computer Science 76), B. Childs, M. Scott, J. Daniel, E. Denman, and P. Nelson, eds., Springer-Verlag, New York, pp. 67–68.
- [10] Ozisik, K. M., 1973, *Radiative Transfer and Interaction With Conduction and Convection*, John Wiley and Sons, New York, pp. 521–522.
- [11] Eckert, E. R. G., and Drake, R. M., 1972, *Analysis of Heat and Mass Transfer*, McGraw-Hill Kogakusha LTD, Tokyo, p. 308.
- [12] Malpica, F., Campo, A., and Tremante, A., 1986, "Contribution of Thermal Radiation to the Temperature Profile of Semitransparent Materials," *High Temp.-High Press.*, **18**, pp. 35–41.
- [13] Siegel, R., and Spuckler, C. M., 1994, "Approximate Solution Methods for Spectral Radiative Transfer in High Refractive Index Layers," *Int. J. Heat Mass Transf.*, **17**, pp. 403–413.
- [14] Patankar, S. V., 1980, *Numerical Heat Transfer and Fluid Flow*, Hemisphere, Washington DC, pp. 83–84.

Gas Turbine Engine Durability Impacts of High Fuel-Air Ratio Combustors—Part I: Potential for Secondary Combustion of Partially Reacted Fuel

S. P. Lukachko

D. R. Kirk

I. A. Waitz

Gas Turbine Laboratory,
Department of Aeronautics and Astronautics,
Massachusetts Institute of Technology,
Cambridge, MA 02139

Demand for greater engine efficiency and thrust-to-weight ratio has driven the production of aircraft engines with higher core temperatures and pressures. Such engines operate at higher fuel-air ratios, resulting in the potential for significant heat release through the turbine if species, such as CO and HC, are emitted from the combustor in large quantities. This paper outlines the magnitude and potential for turbine heat release in current and future engines. The analysis suggests that high fuel-air ratio designs may have to consider changes to cooling strategies to accommodate secondary combustion. A characteristic time methodology is developed to evaluate the chemical and fluid mechanical conditions that lead to combustion within the turbine. Local species concentrations partly determine the potential for energy release. An energy release parameter, here defined as a maximum increase in total temperature (ΔT_t), is used to specify an upper limit on the magnitude of impact. The likelihood of such impacts is set by the convective, mixing, and chemical processes that determine the fate and transport of species through the turbine. Appropriately defined Damköhler numbers (Da)—the comparative ratio of a characteristic flow time (τ_{flow}) to a characteristic chemical time (τ_{chem})—are employed to capture the macroscopic physical features controlling the flow-chemistry interactions that lead to heat release in the turbine. [DOI: 10.1115/1.1584479]

1 Introduction

Demand for greater engine efficiency and thrust-to-weight ratio has driven the production of aircraft engines with higher core temperatures and pressures. Such engines operate at higher fuel-air ratios, resulting in the potential for significant heat release through the turbine if species such as CO and HC are emitted from the combustor in sufficient quantities. Figure 1 illustrates the processes that lead to continuing heat release in the turbine, referred to herein as *secondary combustion*. For fuel-air ratios approaching stoichiometric, increased combustor exit concentrations of CO, OH, H_2 , and O are thermodynamically favored. Falling temperatures and pressures along the turbine flow path compel these species to recombine. This can raise the effective temperature of the turbine flow, impacting performance, [1,2], and adversely affecting durability through augmented heat transfer to internal surfaces. Perhaps more importantly, unmixedness and inadequate residence time will raise concentrations of key species, CO and HC in particular, well beyond the equilibrium limit. If these species are convected close to turbine surfaces, they may interact with oxidizer injected into the gas path in the form of cooling or purge flows, leading to elevated heat flux to cooled parts.

This paper outlines the magnitude and potential for turbine heat release in current and future engines. Particular emphasis is given to interactions with cooling flows. Section 2 discusses the definition and use of fluid and chemical time scales to establish the basis for the analyses presented in this paper. Section 3 details the

key thermodynamic, chemical, and fluid drivers of turbine heat release, including the origins of increased emissions of relevant species into the turbine. Turbine flow compositions representative of current and future engines are defined. In Section 4, time scales and example parameters typical of modern turbines are evaluated to determine the conditions and processes that are important to the mode and scope of turbine heat release. The analysis outlined in this paper demonstrates that turbine heat release may be a common occurrence, even in current engines, with the distinction that higher fuel-air ratio systems are associated with an increased potential for significant local temperature rise on the scale of a turbine blade row. Combined with experimental heat transfer results discussed in a companion paper, [3], these results indicate that higher fuel-air ratio designs may have to consider changes to cooling strategies to accommodate heat release resulting from secondary combustion.

2 Important Physical Parameters

The likelihood and impact of turbine heat release rely on the convective, mixing, and chemical processes that determine the fate and transport of species such as CO and HC. Appropriately defined Damköhler numbers (Da)—the comparative ratio of a characteristic flow time (τ_{flow}) to a characteristic chemical time (τ_{chem})—are employed to capture the macroscopic, physical features controlling the associated flow-chemistry interactions. In this section, important flow, chemical, and thermodynamic parameters are discussed as the basis for specifying Da.

Since we are primarily interested in assessing the impacts on turbine durability, we define a convective flow time scale, τ_{conv} , related to a component length scale, such as a blade chord. Mixing time scales, τ_{mix} , associated with both turbulent diffusion in the freestream and decay of film cooling effectiveness, are also used. Mixing can play an important mitigating role; for example, by

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Amsterdam, The Netherlands, June 3–6, 2002; Paper No. 2002-GT-30077. Manuscript received by IGTI, December 2001, final revision, March 2002. Associate Editor: E. Benvenuti.

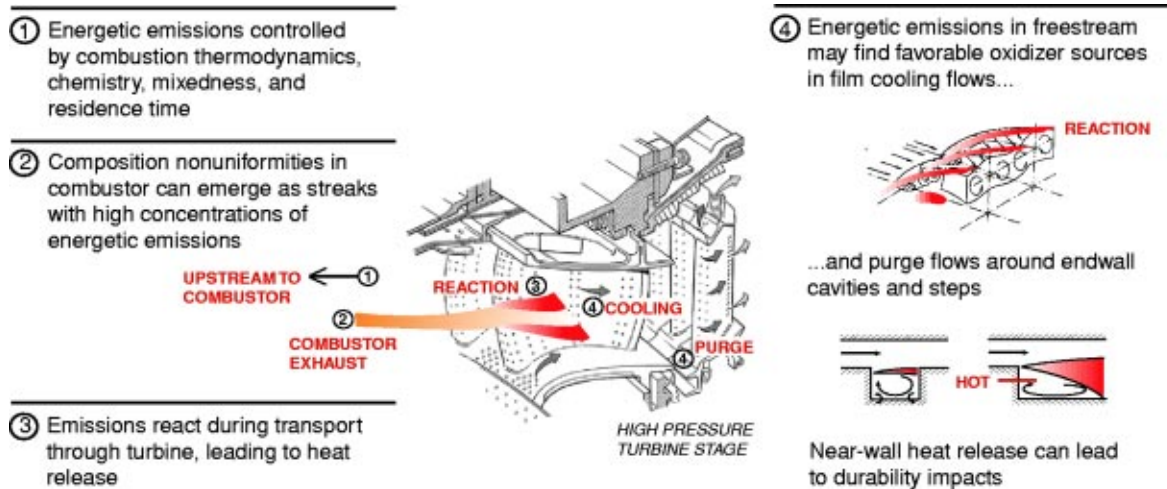


Fig. 1 Phenomenology of secondary combustion in the turbine. (Original engine and film cooling graphics from [19] and [20].)

diffusing streaks of reactive species through the turbine and through limiting the rate at which these species combine with oxygen rich film cooling flows.

The relevant chemical time scales are the time required for a given mixture to ignite (τ_{ign}) and the time required to complete the reaction process (τ_{react}). To the limits of the chemical mechanism used, [4], chemical time definitions are intended to provide an estimate of the fastest reaction rates possible for an estimated turbine flow mixture composition. Chemical times are determined using constant pressure, adiabatic, chemical kinetic calculations for a homogenous flow with no mixing, [5]. Energy released during reaction is assumed to result in temperature rise with no contribution to alterations in the flow kinetic energy. These calculations are initiated at the highest local temperature encountered during the flow interaction hypothesized, therefore assuming that the hot-side temperature of a mixing layer drives ignition and reaction, [6]. This means, for example, that in situations where the core flow encounters a colder flow, such as film cooling, we use the freestream temperature to initiate the calculation.

Local species concentrations partly determine the potential for energy release. An energy release parameter, here defined as a maximum increase in total temperature (ΔT_t), is used to specify an upper limit on the magnitude of impact. The chemical time scales τ_{ign} and τ_{react} are specified explicitly in reference to ΔT_t . For this study, ignition is defined as the time to reach 5% of ΔT_t and the completion of reaction is defined as the time to reach 95% of ΔT_t . The value of ΔT_t , determined by a flame temperature relevant to a local, well-mixed, hypothetically isolated mixture, bounds the magnitude of heat release. As such, in this limit, we do not consider the mitigating impact of mixing on maximum temperature rise. By local we refer to a geometric scale that is much less than the one of interest. For instance, the local mixture could be that contained in a spherical volume of the scale of a cooling hole, the diameter of which is small relative to a turbine blade chord.

There are two important sets of inputs required for the definition of ΔT_t . The chemical composition of the turbine flow involved in a heat release event determines the available energy. This requires a knowledge of combustor emissions speciation. So as to be pertinent to potential secondary combustion, the available energy is specified with reference to the lower heating value (LHV), again an upper limit. Thermodynamically, only part of this available energy will result in ΔT_t . Local temperature and pressure are used to specify the thermodynamic state of the mixture prior to combustion. The turbine operating point and aerodynamics determine the relevant temperatures and pressures.

The above parameters lend themselves directly to three definitions of Da . The magnitude of Da is indicative of the potential for and character of an event. In general, cases where $Da \ll 1$ are limited by reaction and cases where $Da \gg 1$ are limited by flow phenomena. The first definition of Da is the ratio of convective to ignition time, $\tau_{\text{conv}}/\tau_{\text{ign}}$, where τ_{conv} is calculated relative to a blade chord. If the ratio is much less than 1, the potential for any heat release to occur within the scale of a blade is minimal. In essence, the heat release is limited by chemical reaction and locally, the opportunity for surface damage is small. To understand whether heat release can occur over a length scale relevant for turbine durability concerns, the ratio of convective to reaction time, $\tau_{\text{conv}}/\tau_{\text{react}}$, is important. If the ratio is approximately 1, heat release will occur in a distributed manner and the maximum local temperature increase may not be seen. In any case, the impact on heat transfer (increased local heat flux) is related to the change in local surface temperature, but mixing has a key role in determining the extent of impact. The final definition of Da employed, the ratio of mixing to reaction time, $\tau_{\text{mix}}/\tau_{\text{react}}$, is useful for characterizing the distance to which reactive emissions can reach through the turbine since nonuniformities at the combustor exit may mix relatively slowly. Similarly, mixing interactions with cooling jets determine the location of the temperature rise relative to the surface.

3 Emissions Into the Turbine

This section discusses the thermodynamic and practical drivers that determine concentrations of species with potential for secondary heat release at the combustor exit. Section 3.1 outlines the limits of combustor efficiency defined by thermodynamic considerations. Section 3.2 looks at the important role of combustor mixing in determining the level of energy entering the turbine. Finally, Section 3.3 outlines three turbine flow compositions that are representative of current and future engines. These mixture definitions are used in subsequent sections to assess the magnitude and potential for turbine heat release.

3.1 Thermodynamic Efficiency Limits for High Fuel-Air Ratio Combustors. As temperatures and pressures increase, higher concentrations of CO, OH, H_2 , O, and other dissociated species, are found at thermodynamic equilibrium. Figure 2 shows equilibrium concentrations of CO, OH, H_2 , and O for aviation fuel, modeled as $C_{12}H_{23}$, combusted in air, [7]. Two conditions are shown in Fig. 2, one characteristic of a 1970s era commercial aircraft engine at the cruise condition ($T_{T3}=545$ K, $P_{T3}=3.25$ atm) and another that might be characteristic of a 2010s

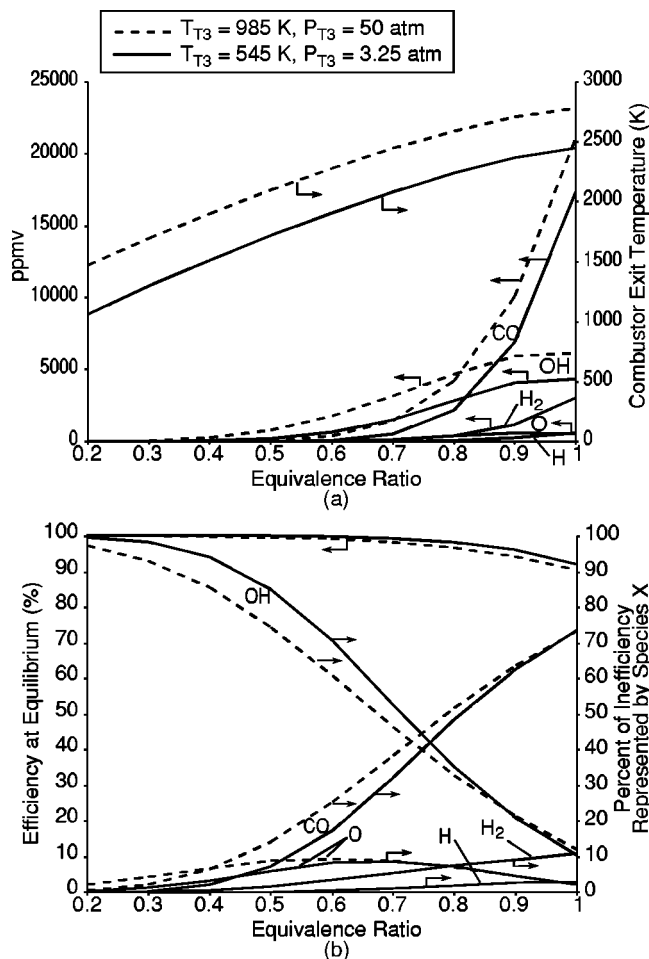


Fig. 2 X_{CO} , H_2 , X_{OH} , X_O , and efficiency (η_b) at constant enthalpy and pressure equilibrium for lean and stoichiometric aviation fuel-air mixtures

era military aircraft engine at takeoff ($T_{T3} = 985 \text{ K}$, $P_{T3} = 50 \text{ atm}$). Combustor exit temperatures resulting from a range of fuel-air ratio specifications are also shown. A lean overall equivalence ratio (ϕ) of 0.3 is typical of current era commercial combustors. Future military applications may demand operation at ϕ near 1. Also shown in Fig. 2 is a measure of combustor efficiency (η_b), defined with reference to lower heating values (LHV). The relative contributions of species that represent the inefficiency ($1 - \eta_b$) are indicated.

Importantly, where idle has historically been the most inefficient condition, future engines will also contend with inefficiency at high power conditions. As temperature increases for a constant pressure, η_b decreases and a greater proportion of the fuel energy enters the turbine gas path. Note that not all of the equilibrium inefficiency exhibited at higher temperatures is contained in the measurable emissions of CO. Other dissociated species, particularly OH, H_2 , and O, are important components of inefficiency at high temperatures. For $\phi = 1$, a significant fraction of the fuel energy, a maximum of $\sim 10\%$ for the cases shown, may be available for heat release in the turbine, more if combustion is not complete. Reducing P_{T3} from the values shown in Fig. 2 would lead to a further decline in efficiency resulting from increases in equilibrium levels of dissociated species for a given turbine inlet temperature (TIT). At equilibrium, as equivalence ratios increase, CO becomes the primary source of inefficiency, replacing OH. However, as the next section will discuss, unmixedness in practical devices promotes species where oxidation is kinetically lim-

ited, such as CO and HC, to levels much greater than equilibrium such that they can become the dominant carriers of energy into the turbine.

3.2 Impacts of Combustor Mixedness and Residence Time.

Historical emissions measurements indicate that combustor efficiencies for current commercial and military aircraft engines at high power conditions are typically above 99.9%, [8–11]. The levels of CO and HC concentrations associated with these efficiencies are several times greater than equilibrium and inefficiencies can be higher for low power conditions such as idle. Deviations from equilibrium are due to the coupled effects of inadequate combustor residence time, less than perfectly mixed fuel and air, requirements for liner cooling (manifest in the profile factor), and tailoring of the TIT for turbine life. Current engines exhibit higher efficiency at higher power conditions and there is a general trend towards lower fractional energy in HC as power level (and thus engine temperature and pressure) is increased, [8,9], primarily a result of better mixing.

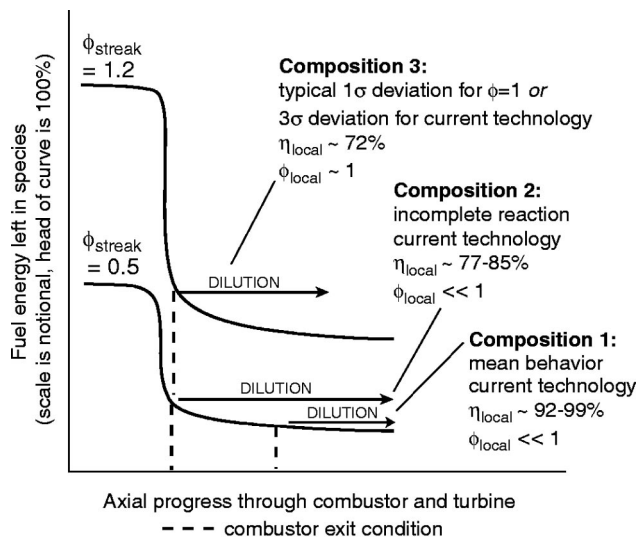
Given the larger equilibrium levels of emissions with potential for secondary heat release, higher fuel-air ratio designs will place a premium on improved mixedness to avoid turbine durability and performance problems. While there is a proportional decrease in mixedness with an increase in primary zone equivalence ratio for the same geometry, [12,13], history suggests that better dilution hole placement, fuel injection, air or fuel staging, and improved liner cooling have made it possible to increase engine temperatures while maintaining combustor efficiency. However, as ϕ approaches stoichiometric levels, near perfect mixing (and thus limited temperature profile tailoring) is required to maintain combustor efficiency near the equilibrium limit.

One important characterization of mixedness is found in the distribution of fuel-air ratio at the combustor exit. As fuel-air ratios increase and efficiencies decrease, there is a greater probability of finding a rich mixture in the turbine. High equivalence ratio nonuniformities emitted into the turbine can be regarded as streaks, and the combustor exit may be characterized by several small-scale or fewer large-scale nonuniformities. Such rich mixtures are more likely to flow further through the turbine than a well-stirred, leaner mixture because of the addition of a mixing constraint beyond the chemistry required for oxidation. This increases the likelihood that heat release may impact blade surfaces.

3.3 Representative Turbine Flow Compositions.

The complexity of combustor flow fields suggests that a wide range of turbine flow compositions is possible. In this section, we develop three representative cases where local speciation in the turbine is estimated for the purpose of determining the likelihood and potential magnitude of secondary combustion. For the estimate we assume that the combustor exit composition is a good approximation for mixtures that might be found through the turbine, with the exception that local mixing with cooling or purge air may perturb the mixture composition. The Appendix discusses the validity of this assumption through an analysis of the mixing of rich streaks emitted by the combustor. It is found that over the length of a turbine stage, reactive streaks larger than $\sim 20\%$ of span typically do not have time to fully mix with the surrounding flow.

Turbine flow compositions are constructed by tracking the fate of a streak of fluid, with an assumed fuel type and initial equivalence ratio (ϕ_{streak}) that reflects the state of mixedness in the combustor. To estimate the post-combustor composition, equilibrium is calculated for a specified temperature and pressure relevant to the local site, then changes are made to the resulting concentrations of CO, HC, and H_2 to specify a relevant range of mixing ratios for these kinetically limited species. Compositions of OH, O, and other species are assumed to be at their equilibrium levels and thus, their mole or mass fractions are byproducts of the above specifications. Because it is less likely that high carbon number species would emerge from the combustor, all HC was assumed to be CH_4 , chosen also for its inhibitory impact on CO and H_2



Composition	CO-equivalent (ppmv)
1	1400 - 12000
2	21000 - 32000
3	77000 - 81500

Fig. 3 Representative current and future turbine flow compositions. (Portion of fuel energy represented by CO, HC, and/or H₂ specified at 99%, 85%, and 73% for compositions 1 through 3. η_{local} reduced with changes in local T and P .)

oxidation and thus to specify a lower bound for kinetic rates. Because we are primarily concerned with durability impacts, we additionally perturb the mixture to account for mixing with oxidizer associated with local cooling or purge flows to provide a worstcase near-surface equivalence ratio (ϕ_{local}). This process and the range of turbine flow conditions that results are illustrated in Fig. 3. Since the total available energy defines the maximum local temperature rise, CO, OH, H₂, O, and HC emissions are summed into a CO-equivalent molar quantity by energy content in Fig. 3.

In Fig. 4, the fuel-lean mixture, Composition 1, with $\phi_{streak} = 0.5$ and $\eta_{local} \sim 92-99\%$, is chosen to represent the mean performance of a well-mixed, current era combustor of high efficiency. The range of efficiency reflects the influence of different combustor exit temperatures and pressures on equilibrium concentrations of species other than CO, HC, and H₂. Composition 2 reflects an increase in CO-equivalent emissions for the same current era combustor as described for Composition 1, resulting in $\eta_{local} \sim 77-85\%$. This decline in efficiency could, for example, be due to lack of adequate residence time for complete combustion. The fuel-rich mixture, Composition 3, with $\phi_{streak} = 1.2$ and $\eta_{local} \sim 72\%$, represents the additional dimension of unmixedness. This case selects, for example, the 1 σ point along the upper tail of a future, stoichiometric combustor with an $S = \sigma/\phi = 0.2$ Gaussian exit distribution ([14], see also [15]), a level of unmixedness typical of current combustors. Composition 3 is also consistent with a greater than 3 σ deviation along the upper tail of the Composition 2 case. Thus, Composition 3 is more likely for future turbine flows as compared to current designs. The resulting mole fraction of emissions that may be involved in heat release for Composition 3 is in excess of 75,000 ppmv equivalent CO by energy, nearly three times the stoichiometric equilibrium value. Six fractional distributions of CO, HC, and H₂ were employed for each composition specified to assess the kinetic impact of different assumptions for partitioning among these species, which are expected to represent the bulk of the combustion inefficiency entering the turbine. For

cases that include hydrogen, a 10% fractional distribution was assumed. The HC/CO fraction was varied from 0.5 to 1.0 for each case where both were included in the mixture.

4 Oxidation of Emissions in the Turbine

In the previous section, we proposed a reasonable range of estimates for the quantity of species entering the turbine with potential for secondary heat release for current and future engines. In the high-pressure turbine (HPT), oxidative chemical interactions occur among air constituents, primary combustion products, and trace species (e.g., CO, H₂, NO_y, SO_x, HO_x, and O), [16,17]. These oxidative processes are quenched by rapid fluid property changes through the HPT. Thus, whether these species react within the turbine is a function of the time evolution of temperature and pressure through the turbine, local fluid mechanical effects, as well as the constituency of such emissions.

Section 4.1 places bounds on the magnitude of energy release that can occur as a result of such emissions by estimating ΔT_T for the composition scenarios described in the previous section. However, only part of the energy resident in emissions from the combustor is recovered through the turbine. In Section 4.2, we consider the potential heat release by estimating Da. In the course of this discussion, we also discuss the importance of various factors that govern chemical time scales. For illustrative purposes, we investigate ΔT_T and Da bounds over a temperature and pressure range of 900–2500 K and 0.1–45 atm, respectively, representing the local temperatures and pressures that might be experienced within the engine over a range of operating conditions. A subset of temperatures and pressures, between approximately 1500–2100 K and approximately 10–25 atm, is employed as characteristic of the nozzle guide vane (NGV) flow through a current era advanced engine turbine.

The impact of secondary heat release on durability is dependent on the magnitude of the energy release and subsequently, the augmentation in heat flux that might be experienced. Mixing processes within the freestream and the interaction with oxidizer-rich cooling or purge flows can impact the potential and magnitude of heat release and its impact on local heat release. Section 4.3 looks at mixing of freestream and local oxidizer flows to evaluate potential limitations on surface heat transfer.

4.1 Bounds on the Magnitude of Energy Release. The adiabatic flame temperature is a thermodynamic limit for the magnitude of heat released and subsequent temperature rise resulting from oxidation and recombination in the turbine. Figure 4 shows results for ΔT_T calculated in this manner for the representative turbine conditions described in Fig. 3. Thermodynamic data are taken from [4]. Also shown in Fig. 4 are ranges for H^* , the ratio of enthalpy change to initial enthalpy ($\Delta H_i/H_i$). This is an alternative interpretation of the temperature rise.

The primary driver of ΔT_T is the local concentration of species with the potential for further heat release. Thermodynamically, there is a generally greater dependence on initial local temperature as opposed to pressure, but pressure dependencies are important up to $P \sim 10$ atm. Although there are some effects of composition due to differences in the assumed mixing between freestream and oxidizer, such details are not as important. The local temperature rise, ΔT_T , is greater for low initial temperature and high-pressure conditions. Sensitivity to local conditions increases with greater concentrations of CO and H₂ because their equilibrium values vary more widely over the temperature and pressure space.

Composition 3 (Case C3) exemplifies an upper bound for a high equivalence ratio, future engine of thermodynamically limited efficiency, and unmixedness typical of current era combustors. Over a range of temperatures and pressures characteristic of the NGV, local temperature rise can exceed 500 K, or an H^* of $\sim 35\%$. This is much higher than the approximately 10 K rise ($H^* \sim 0.45\%$) of case C1 for the same temperature and pressure space, which exemplifies a current, highly-efficient, well-mixed engine.

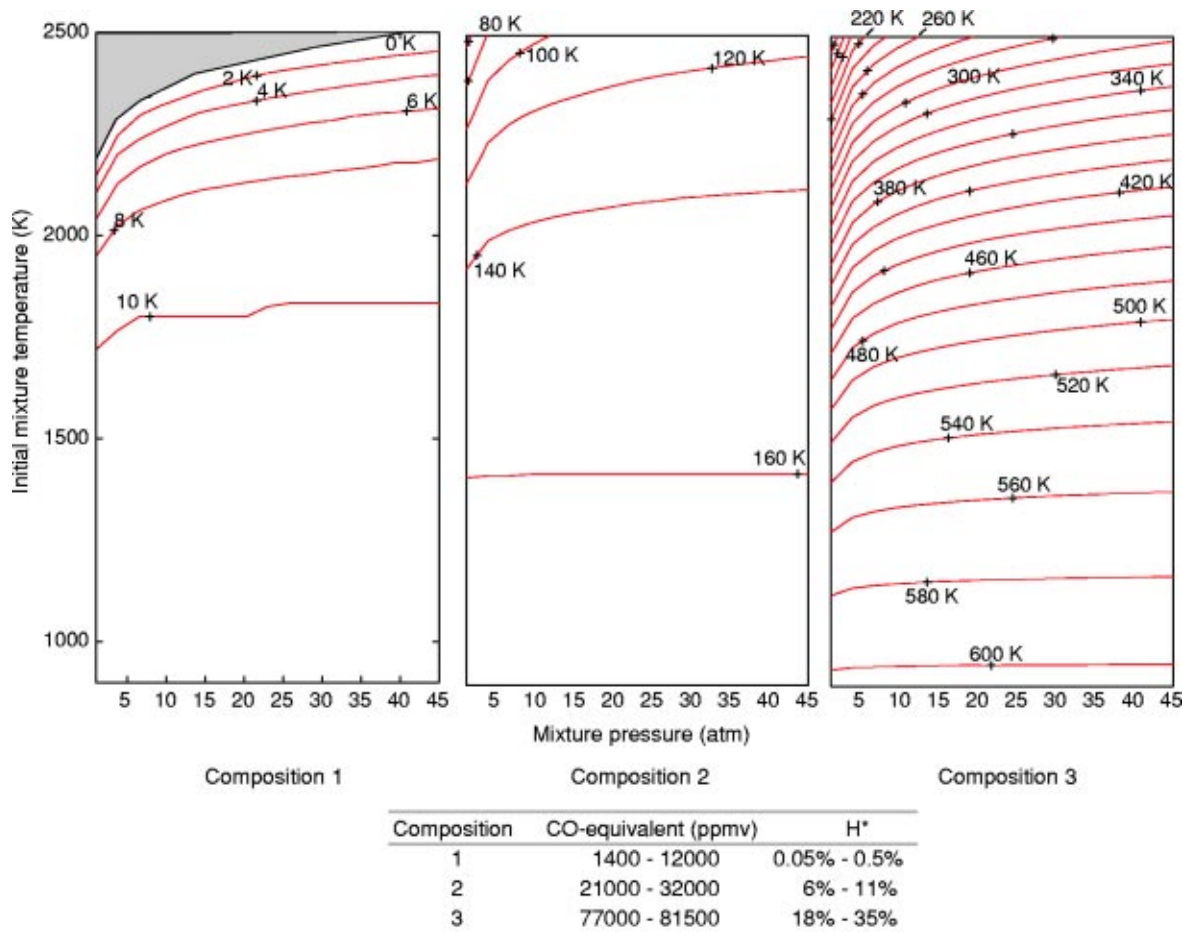


Fig. 4 Potential local temperature rise. (Grayed area indicates thermodynamically incompatible T , P , X combinations.)

Increasing unmixedness for current engines, as exemplified by Case C2, increases the potential local temperature rise to 150 K ($H^* \sim 11\%$). Generally, a combustor efficiency of less than 98.5% is required to realize a temperature rise (ΔT_T) greater than 50 K for Case C3 where an efficiency of less than 95% is required for Case C1.

4.2 Energy Release Depends on Fluid and Chemical Time Scales. Faster reaction rates are associated with a greater risk that heat release will occur locally. To determine the important factors governing chemical time scales, knowledge of the speciation at the combustor exit is required. Because empirical characterizations across the range of potential combustor exit conditions are limited, the turbine flow composition cases discussed previously were used to estimate characteristic chemical time scales. Chemical times— τ_{ign} and τ_{reac} defined by the time to reach 5% and 95% of ΔT_T , respectively—were estimated using a constant pressure, adiabatic, chemical kinetic model for homogenous flow with no mixing using the GRIMech2.1 reaction set, [4]. The method of determining characteristic chemical times was chosen to provide the shortest time and thus a worstcase estimate. Figure 5 plots results for τ_{ign} and τ_{reac} . The boxes in Fig. 5 indicate ranges of median chemical time, with and without H_2 , for a temperature and pressure range of approximately 1500–2100 K and approximately 10–25 atm characteristic of the NGV flow through a current-era advanced engine turbine. The bracketed lines indicate the full range of chemical times for this temperature and pressure range.

In the absence of mixing effects (e.g., strain, turbulence), τ_{ign} is typically on the order of 0.01 ms (0.005 to 0.1 ms range). The

reaction time, τ_{reac} , is typically on the order of 0.1 ms (0.01 to 0.5 ms range). Since the temperatures are high in advanced engines, τ_{ign} is predominantly the time required for requisite radical concentrations to be generated. For the initial composition of the mixtures used for the analysis, radical concentrations are set to the levels found at chemical equilibrium at the given temperature and pressure. For higher equivalence ratio engines, radical concentrations emitted from the combustor may be greater than those assumed and subsequently, τ_{ign} may be lower.

Chemical times are overwhelmingly driven by local temperature, but for any given temperature and pressure, the range of τ_{reac} is less than an order of magnitude for the cases explored, despite the large range in turbine flow composition. For chemistry in the turbine, local reactions occur under significantly more dilute conditions than in the combustor, which may dampen the impact of concentration changes on reaction rate as compared to temperature. In Fig. 5, the impact of composition is smaller than that of temperature with the most consistent perturbation related to the presence of H_2 in the mixture, which affects τ_{ign} most prominently. As equivalence ratios increase, if mixing cannot be improved, HC emissions will increase which can inhibit CO and H_2 oxidation. The inhibitory impact of CH_4 on chemical times is only evident in Case C3, where the local fuel-air ratio is consistently set to a stoichiometric value. For the other composition cases, there is an increase in chemical rates with increased HC fraction, which arises because the substitution of CH_4 for CO effectively increases the equivalence ratio of the lean mixture (see Fig. 3). The nonlinear balance between the total mass of emissions and how fast they are consumed through oxidation contributes to the decline and then increase in characteristic chemical times as the

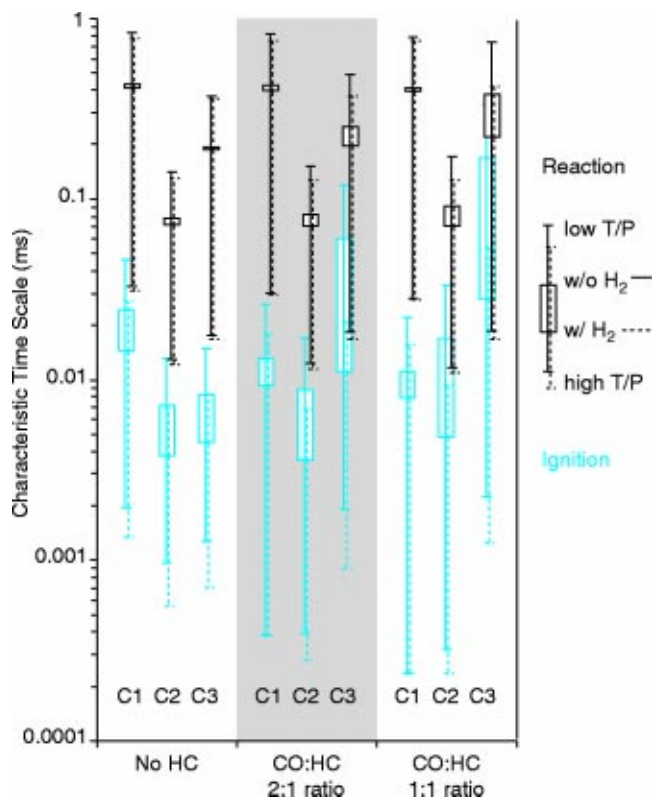


Fig. 5 Characteristic chemical time scales for turbine heat release for representative flow compositions. (Boxes represent median values and bracketed lines indicate ranges for a typical NGV flow space for a current era advanced engine.)

total fraction of emissions increases (Case C1→Case C3). Generally second-order reactions are involved which have rates that are geometric with species concentrations so that for a given ϕ_{streak} , chemical times will decrease with increasing fuel content. However, there is more fuel to consume and a larger relative radical concentration to develop prior to oxidation, trends that act to increase chemical times.

In Fig. 6, a Damköhler number (Da) comparing a convective time of 0.5 ms, typical of a turbine blade row, and the ignition time are shown for the mixture compositions outlined in Fig. 3. The results are overlapped upon ΔT_T results from Fig. 4. Note that this specification of Da essentially decouples the chemical and fluid scales (e.g., there is no dependence on temperature change in τ_{flow}). For the temperatures shown, CO kinetics are explosive over a greater range of pressures as temperatures increase, but lower pressures exhibit a significant change in rates as the explosive regime is exited, shown by the knee in the Da curves at lower temperatures.

Moving from the lower left to the upper right of each temperature and pressure space, the current direction of technology development, conditions for secondary combustion become more favorable because temperatures and pressures (and concentrations) are increasing. Since temperatures are highest at high power conditions, takeoff, climbout, or other accelerations are the most probable time for oxidation and heat release within the turbine. For a current era engine, in which the NGV flow may be described by a temperature and pressure space of 1500–2100 K and 10–25 atm, Da based on a typical blade row convective time of 0.5 ms is 10–500 for ignition and 1–25 for reaction for all cases. For situations in which residence times would be much greater, for recirculation zones for example, the entire temperature and pressure space would consist of high Da for the well-mixed, adiabatic limit.

Given the magnitude of Da over the NGV temperature and pressure space, oxidation of species such as CO due to interactions with cooling air can be fairly common even in current era

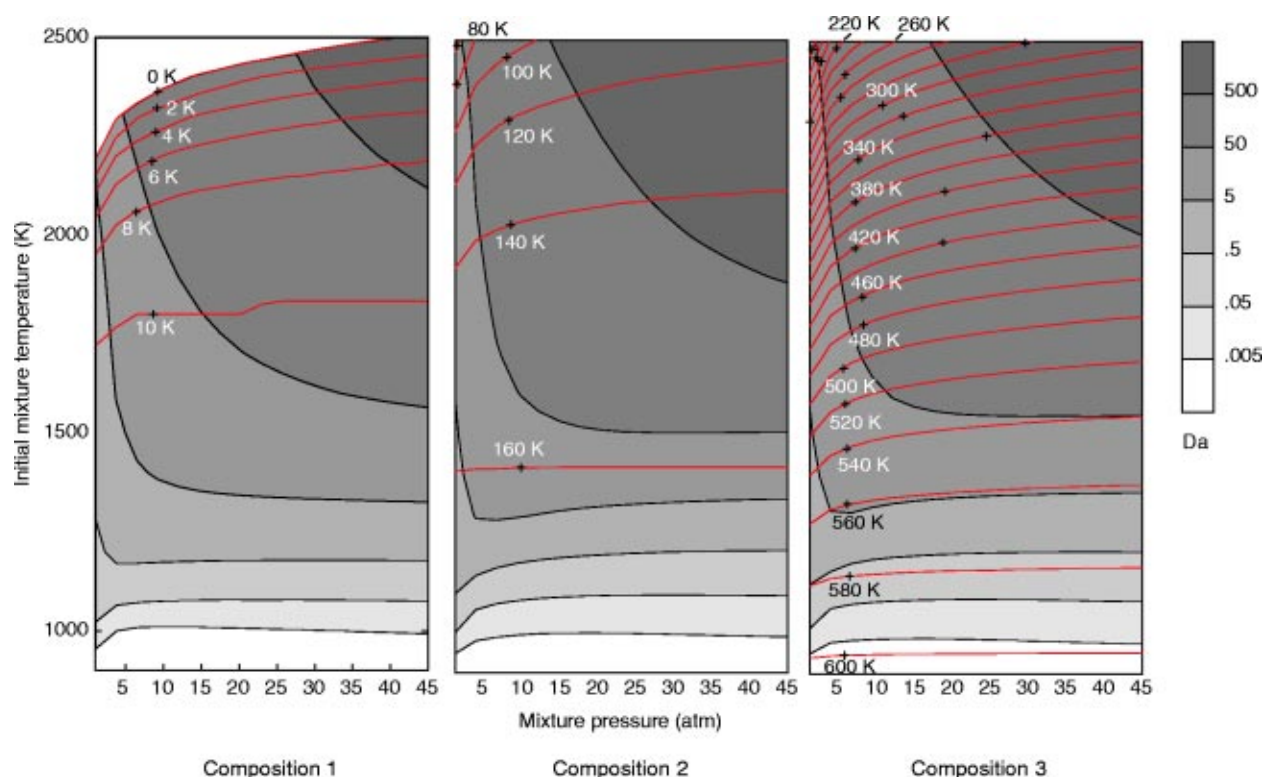


Fig. 6 ΔT_T and Da comparing ignition and blade row traverse times for representative flow compositions. (Blank area indicates thermodynamically incompatible T , P , X combinations.)

turbines, provided fluid mechanical mechanisms of extinction are not in evidence. Within a specific engine, the potential for secondary combustion becomes weaker as the high and low turbines are traversed and temperatures and pressures decline. Following a streamline through the NGV, Da for reaction declines to order one, indicating that the HPT would be the section of most concern for heat release impacts. Furthermore, the magnitude of Da indicates that reactions, if begun in the NGV, may lead to the highest temperatures towards the end of the NGV or beginning of the first rotor of the HPT. For Case C3, which represents an unmixed slug emitted into the turbine of a stoichiometric engine, realizing the full extent of reaction can have a significant impact on durability since maximum temperature increases may exceed 500 K.

4.3 Mixing Mitigates Wall Impacts. Reactions and heat release important to durability will occur within mixing layers between freestream and coolant flows. The rapidity with which near-wall thermal layers change subsequent to coolant injection is a relevant mixing scale that has been well documented (e.g., [18]). In application to the reacting coolant layer durability problem, we calculate a characteristic mixing time scale based upon the decay of adiabatic film effectiveness, η , as a surrogate for species mixing. As a time scale, this analysis essentially provides an estimate of when the cooling oxidizer would be exhausted and a maximum surface impact can be expected. The freestream and coolant are assumed to be initially separated (freestream equivalence ratio $\sim \infty$), subsequently mixing turbulently with $Le=Pr=Sc=1.0$. Since the cooling effectiveness maps the near-wall mixing state and is a nondimensionalized scalar quantity, we replace temperatures with mass fractions (Eq. (1)). For the analysis, we assume that the oxidizer stream is air and that the freestream is a diluted fuel where the O_2 has been exhausted.

$$\eta = (T_{aw} - T_\infty) / (T_c - T_\infty)$$

$$\eta = (Y_{fw} - Y_{f\infty}) / (Y_{fw,i} - Y_{f\infty}) \quad \text{where } Y_{fw,i} = 0 \text{ and } Y_{f\infty} = 1 \quad (1)$$

At the wall, since a bimodal component composition has been assumed, the oxidizer and freestream fractions are directly related. The value of the effectiveness, which indicates a particular mixture of freestream and coolant at the wall, is related to a fuel-air ratio at the wall (FAR_w) which is a function of composition and concentration (Eq. (2)).

$$\begin{aligned} Y_{ow} &= 1 - Y_{fw} \\ FAR_w &= Y_{fw} / Y_{ow} = Y_{f\infty} (1 - \eta) / (1 - Y_{f\infty} (1 - \eta)) \\ \eta &= 1 / (FAR_w + 1) \end{aligned} \quad (2)$$

For threshold η , set $FAR_w = FAR_{\text{thresh}}(Y_{O_2|o} / Y_{(HC+CO)|f})$.

Here, FAR_w is set to specify a stoichiometric wall value to determine when all of the oxidizer would be mixed with the freestream.

For a set FAR_w , the threshold value of η is smaller for increased HC content, giving a longer mixing time scale. This is because the required O_2 for a combustible mixture changes as the fuel composition changes and the relative mass of fuel in freestream increases with CO content. For low concentrations of freestream species with the potential for further heat release, such as Case C1 in Fig. 3, the threshold value of η is ~ 0.01 , increasing to ~ 0.15 for Case C3, with higher freestream CO concentrations. This indicates a significant mixing limitation on the depth into a cooling layer the maximum temperature rise can penetrate. For a typical cooling hole with an attached jet (e.g., [18]), the threshold value of η is not reached along the centerline for distances downstream of greater than 40 cooling hole diameters (D), a scale equivalent to a turbine blade chord. Thermal perturbation of the near-wall flow would be effectively delayed, even in cases where reaction times are fast. In contrast, a lifted coolant jet could realize threshold η values at less than 10^*D which indicates that maximum temperatures could be experienced at more vulnerable

locations between cooling holes. Furthermore, cooling flows are three-dimensional and for typical lateral hole spacings (typically 3^*D), there will be regions where the maximum temperature is reached nearer to the surface than would be supposed by centerline or average effectiveness profiles. However, in reference to Fig. 6, because nonreactive, cooler air is diffused into the reacting zone, the realized temperature rise would be smaller than suggested by the local adiabatic flame temperature. If the resulting mixture temperature is low, coolant flows could quite possibly quench reactions.

5 Conclusion

In this paper, a characteristic time methodology was developed to evaluate the chemical and fluid mechanical conditions that can lead to secondary combustion within aircraft turbines. Representative turbine flow compositions for current and future engines were estimated and used to determine the maximum increase in total temperature (ΔT_t) and therefore specify an upper limit on the magnitude of impact. These compositions were further employed to evaluate the likelihood of such impacts through the definition of Damköhler numbers (Da) appropriate to the macroscopic, physical features controlling the flow-chemistry interactions that lead to heat release in the turbine.

As temperatures and pressures increase, higher concentrations of CO, OH, H_2 , O, and other dissociated species are found at thermodynamic equilibrium. As a result, 10% of the fuel energy may be available for secondary combustion, more as unmixedness in the combustor increases or if the geometry of the combustor is such that there is inadequate residence time in which to complete combustion. Thus, where idle has historically been the most inefficient condition, future aircraft engines may also contend with inefficiency at high power conditions.

Compared to a typical blade row convective time of 0.5 ms, Da based on ignition time is on the order of 10–500 for NGV flow conditions applicable to current era engines over a wide range of species conditions. Comparatively, Da based on reaction time is on the order of 1–25, and reactions, if begun at the end of the blade row, may continue into the next blade row. Thus, provided fluid mechanical mechanisms of extinction are not in evidence, oxidation of species such as CO due to interactions with cooling air can be fairly common even in current era combustors. Since temperatures are highest at high power conditions, takeoff, climbout, or other accelerations are the most probable time for oxidation and heat release within the turbine. Mixing processes work to reduce the extent of reactive, fuel-rich streaks at the combustor exit. However, over the length of a turbine stage, reactive streaks larger than $\sim 20\%$ of span do not have time to fully mix with the surrounding flow. This may be a significant mode in which species with potential for secondary heat release can survive well downstream of the combustor. However, heat release impacts on turbine heat flux are limited by relatively slow mixing with coolant, particularly for attached flows.

Mixedness improvements and other changes applied to bring the combustor exit flow closer to an equilibrium condition are mitigating measures that cannot overcome the fundamental thermodynamic determinants that govern the emission of species with the potential for further heat release into the turbine. In the near term, turbine durability impacts may constrain the introduction of higher temperature and pressure engines. A companion paper, [3], details experiments which measure up to 70% of the temperature rise suggested by ΔT_t for a single row of reacting film cooling jets. As fuel-air ratios increase further, an unprecedented merging of combustor and turbine designs may be required. If future aircraft engines are to continue increases in peak temperature and pressure, combustor and turbine design procedures will need to be altered to accommodate secondary combustion.

Acknowledgments

This effort was sponsored by the generous support of Pratt & Whitney. The authors would like to thank Drs. Frederick Soechting, Charles Graves, Thomas Auxier, Saadat Syed, William Kvasnak, and Atul Kohli of Pratt & Whitney, Dr. Catalin Fotache of the United Technologies Research Center, and Prof. Edward Greitzer of MIT for their guidance and critique of this research.

Nomenclature

ΔH_t	= change in total enthalpy
ΔT_t	= change in total temperature
η	= film effectiveness
η_b	= overall combustor/burner efficiency
η_{exit}	= local efficiency at combustor exit
η_{local}	= turbine flow combustion efficiency local to surface
ϕ	= overall combustor equivalence ratio
ϕ_{exit}	= local equivalence ratio at combustor exit
ϕ_{local}	= turbine flow equivalence ratio local to surface
ϕ_{streak}	= initial streak equivalence ratio
σ	= standard deviation
τ_{chem}	= characteristic chemical time scale
τ_{mix}	= characteristic mixing time scale
τ_{flow}	= characteristic flow time scale
τ_{ign}	= ignition time
τ_{react}	= reaction time
D	= cooling hole diameter
Da	= Damköhler number = $\tau_{\text{flow}}/\tau_{\text{chem}}$
El	= emission index
FAR	= fuel-air ratio
H^*	= enthalpy release ratio = $\Delta H_t/H_t$
H_t	= total enthalpy
LHV	= lower heating value
NGV	= nozzle guide vane
P	= pressure
S	= mixedness parameter = σ/ϕ
T	= temperature
TIT	= turbine inlet temperature
X	= mole fraction
Y	= mass fraction

Subscripts

∞	= freestream
a	= adiabatic
c	= coolant
f	= fuel
i	= initial
o	= oxidizer
$T3$	= combustor inlet
thresh	= threshold
w	= wall

Appendix

Mixing of Fuel Rich Streaks Enroute to Downstream Surfaces. Low mixedness is manifest at the combustor exit by spatial nonuniformities in fuel-air ratio. As a result of low mixedness, and in contrast to the emission of CO or HC resulting from inadequate residence time, some portion of the flow will be characterized by a greater than stoichiometric composition. These rich streaks can transport emissions downstream of the combustor where mixing with oxygen-rich film cooling or purge flows can occur. Streaks that carry lean mixtures will be further diluted at the site and are relatively less consequential, both because of the reduced opportunity for reaction and the diminished impact of the subsequent heat release. It can be expected that upon emission, a rich streak would about other regions of the exit flow that are lean. The fate of the streak—whether a portion of the streak survives the transit—depends on the scale of the nonuniformity as well as the relative mixing, convective, and chemical time scales.

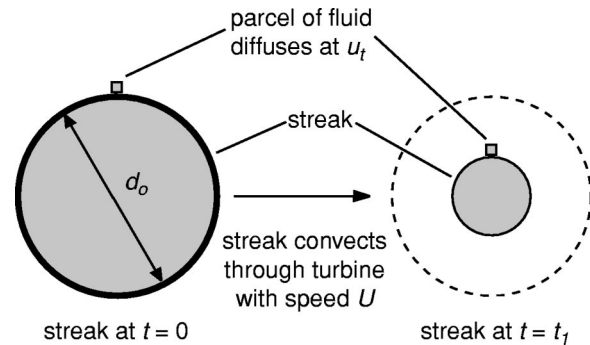


Fig. 7 Turbulent diffusive mixing

We determine the importance of a particular nonuniformity by the definition of a minimum length scale, representing the smallest streak that survives a single, representative HPT stage. Two simplified mixing models were employed to measure the extent of mixing in transit, a diffusive transport analysis governed by turbulence levels, and the development of a shear layer due to a velocity differential. The resulting estimates were used to define $\tau_{\text{mix}}/\tau_{\text{conv}}$. If $\tau_{\text{mix}}/\tau_{\text{conv}}$ is greater than one, some portion of the streak will arrive unperturbed. If oxidation reactions are slow, such that Da , defined by the reaction time and appropriate convective scale, $\tau_{\text{conv}}/\tau_{\text{reac}}$, is small, some of the partially reacted fuel in the mixed-out region may also survive.

Turbulent diffusion is one of the primary mechanisms through which a streak is mixed with adjacent flow. The temperature ratio of the streak to adjacent flow is taken to be unity and thus the density and velocity ratios are also taken to be unity. Under the assumption of isotropic turbulence with no cross velocities, no shear stresses form and thus, no shear layer develops. Additionally, no superimposed molecular diffusion is employed as it is assumed that the turbulence is dominant. In the manner shown in Fig. 7, the diffusion, or mixing time, τ_{mix} , is defined as the time that it takes for a parcel of fluid at the outside of the reactive streak to diffuse to the centerline of the streak.

An estimate for τ_{mix} is dependent on the diffusion velocity (V_d) which is related to the turbulent diffusivity (D_t) by analogy to the equation for Fickian diffusion. For turbulent diffusion, V_d is approximately the turbulent velocity, u_t . The turbulent velocity is estimated from typical levels of turbulent intensity ($u'/U = 0.1-0.15$) at the combustor exit. Using the assumption of homogeneous isotropic turbulence, the turbulent velocity is estimated from the turbulent intensity as approximately u' . This gives a conservative definition for the mixing time, $\tau_{\text{mix}} = d_o/(2u_t)$.

Streak parameters such as temperature, pressure, position, and time were taken from streamlines calculated for a current era advanced engine. Turbulent intensities were varied from 5–50% of the freestream velocity and a range of streak diameters from 5–50% of the combustor exit span was considered. Results indicated that streaks larger than $\sim 10\%$ of span will not have sufficient time to mix out over the length scale of a turbine stage. The same approach can be used on other downstream components.

In addition to the turbulent diffusive mixing, the development of a shear layer between the streak and the adjacent flow also provides a potential mixing mechanism. Different velocities will result from different stream temperatures. Analytical distributions of velocity and pressure for a two-dimensional, turbulent shear layer between two co-flowing streams of different velocities and temperatures were employed for the analysis. Temperature differences were derived from typical combustor exit profiles. The temperature difference leads to the velocity nonuniformity between the cold and hot streams, either of which may carry emissions that may be oxidized. The definition of τ_{mix} in this case is conservatively defined as the time required for the velocity at the centerline

of the streak to slow down by 1%. This is essentially a criterion for the penetration of the shear layer to the center of the streak.

The shear mechanism provides a more vigorous mechanism for mixing than diffusive turbulence. Over the length of a turbine stage, and for a range of velocity ranges, it was found that reactive streaks larger than $\sim 20\%$ of span do not have time to fully mix with the surrounding flow. Thus, in the limit of fast chemistry, $\tau_{\text{conv}}/\tau_{\text{reac}} \gg 1$, some amount of emissions can survive through the HPT unmixed.

References

- [1] Sirignano, W. A., and Liu, F., 1999, "Performance Increases for Gas-Turbine Engines Through Combustion Inside the Turbine," *J. Propul. Power*, **15**, pp. 111–118.
- [2] Godin, Th., Harvey, S., and Stouffs, P., 1997, "Chemically Reactive Flow of Hot Combustion Gases in an Aircraft Turbo-Jet Engine," ASME Paper No. 97-GT-302.
- [3] Kirk, D. R., Guenette, G. R., Lukachko, S. P., and Waitz, I. A., 2002, "Turbine Durability Impacts of High Fuel-Air Ratio Combustors, Part 2: Impact of Intra-Turbine Heat Release on Film-Cooled Surface Heat Transfer," ASME Paper No. GT-2002-30182.
- [4] Bowman, C. T., Hanson, R. K., Davidson, D. F., Gardiner, W. C., Jr., Lissianski, V., Smith, G. P., Golden, D. M., Frenklach, M., and Goldenberg, M., 1995, "GRI-Mech 2.11, http://www.me.berkeley.edu/gri_mech/."
- [5] Kee, R. J., Rupley, F. M., and Miller, J. A., 1991, "CHEMKIN-II: A FORTRAN Chemical Kinetics Package for the Analysis of Gas-Phase Chemical Kinetics," SAND89-8009, Sandia National Laboratories, Livermore, CA.
- [6] Fotache, C. G., Wang, H., and Law, C. K., 1999, "Ignition of Ethane, Propane, and Butane in Counterflow Jets of Cold Fuel Versus Hot Air Under Variable Pressures," *Combust. Flame*, **117**, pp. 777–794.
- [7] Gordon, S., and McBride, B. J., 1994, "Computer Program for Calculation of Complex Chemical Equilibrium Compositions and Applications I. Analysis," NASA-RP-1311, NASA Glenn Research Center, Cleveland, OH.
- [8] International Civil Aviation Organization, 1995, *ICAO Engine Exhaust Emissions Data Bank*, First Ed. With Addendums, ICAO Doc 9646-AN/943. Montreal, Canada.
- [9] U.S. Environmental Protection Agency, 1992, "Procedures for Emission Inventory Preparation: Volume IV, Mobile Sources," EPA420-R-92-009, Washington, DC.
- [10] Wey, C. C., et al., 1998, "Engine Gaseous, Aerosol Precursor and Particulate at Simulated Flight Altitude Conditions," NASA-TM-1998-208509, NASA Glenn Research Center, Cleveland, OH.
- [11] Howard, R. P., et al., 1996, "Experimental Characterization of Gas Turbine Emissions at Simulated Flight Altitude Conditions," AEDC-TR-96-3, Arnold Engineering Development Center, Arnold Air Force Base, TN.
- [12] Sturgess, G. J., McKinney, R., and Morford, S., 1992, "Modification of Combustor Stoichiometry Distribution for Reduced NO_x Emission from Aircraft Engines," ASME Paper No. 92-GT-108.
- [13] Heywood, J. B., and Mikus, T., 1973, "Parameters Controlling Nitric Oxide Emissions from Gas Turbine Combustors," Paper 21, Presented at AGARD Propulsion & Energetics Panels 41st Meeting on Atmospheric Pollution by Aircraft Engines, London, England.
- [14] Mikus, T., and Heywood, J. B., 1971, "The Automotive Gas Turbine and Nitric Oxide Emissions," *Combust. Sci. Technol.*, **4**, pp. 149–158.
- [15] Fric, T. F., 1993, "Effects of Fuel-Air Unmixedness on NO_x Emissions," *Journal of Propulsion and Power*, **9**(5), pp. 708–713.
- [16] Lukachko, S. P., Waitz, I. A., Miake-Lye, R. C., Brown, R. C., and Anderson, M. R., 1998, "Production of Sulfate Aerosol Precursors in the Turbine and Exhaust Nozzle of an Aircraft Engine," *J. Geophys. Res.*, **103**(D13), pp. 16, 159–16, 174.
- [17] Penner, J. E., Lister, D. H., Griggs, D. J., Dokken, D. J., and McFarland, M., eds., 1999, "Special Report on Aviation and the Global Atmosphere," IPCC, WMO/UNEP, Cambridge University Press, Cambridge, UK.
- [18] Harnett, J. P., 1985, "Mass Transfer Cooling," *Handbook of Heat Transfer Fundamentals*, W. M. Rohsenow, J. P. Hartnett, E. N. Ganic, eds., McGraw-Hill, New York, Chap. 1.
- [19] Rolls-Royce, plc., 1992, *The Jet Engine*, 4th Ed., Derby, England.
- [20] Lakshminarayana, B., 1996, *Fluid Dynamics and Heat Transfer of Turbomachinery*, New York.

Gas Turbine Engine Durability Impacts of High Fuel-Air Ratio Combustors—Part II: Near-Wall Reaction Effects on Film-Cooled Heat Transfer

D. R. Kirk

Graduate Research Assistant

G. R. Guenette

Principal Research Engineer

S. P. Lukachko

Research Engineer

I. A. Waitz

Professor of Aeronautics and Astronautics

Gas Turbine Laboratory,
Department of Aeronautics and Astronautics,
Massachusetts Institute of Technology,
Cambridge, MA 02139

As commercial and military aircraft engines approach higher total temperatures and increasing overall fuel-to-air ratios, the potential for significant chemical reactions on a film-cooled surface is enhanced. Currently, there is little basis for understanding the effects on aero-performance and durability due to such secondary reactions. A shock tube experiment was employed to generate short duration, high temperature (1000–2800 K) and pressure (6 atm) flows over a film-cooled flat plate. The test plate contained two sets of 35 deg film cooling holes that could be supplied with different gases, one side using air and the other nitrogen. A mixture of ethylene and argon provided a fuel rich freestream that reacted with the air film resulting in near wall reactions. The relative increase in surface heat flux due to near wall reactions was investigated over a range of fuel levels, momentum blowing ratios (0.5–2.0), and Damköhler numbers (ratio of flow to chemical time scales) from near zero to 30. For high Damköhler numbers, reactions had sufficient time to occur and increased the surface heat flux by 30 percent over the inert cooling side. When these results are appropriately scaled, it is shown that in some situations of interest for gas turbine engine environments significant increases in surface heat flux can be produced due to chemical reactions in the film-cooling layer. It is also shown that the non-dimensional parameters Damköhler number (Da), blowing ratio (B), heat release potential (H^*), and scaled heat flux (Q_s) are the appropriate quantities to predict the augmentation in surface heat flux that arises due to secondary reactions.

[DOI: 10.1115/1.1606473]

1 Introduction

Requirements for increased specific thrust mandate high operating temperatures in modern aircraft engines and consequently higher fuel-to-air ratios in the combustor. As the fuel-to-air ratio of the combustor is increased, there is added risk of reactive species interacting with film-cooled surfaces where further reactions and heat release may occur. The source of reactive species may be combustor unmixedness or, for fuel-air ratios approaching stoichiometric, large quantities of dissociated species, such as CO [1]. The additional heat loads due to reaction must be quantified so that suitable compensation in film-cooling strategies or structural robustness can be made. In this context, an experiment has been designed and implemented with the objective of investigating the change in surface heat flux due to local reactions occurring as a result of the interaction of a fuel-rich freestream and a film-cooling layer. The experimental results were combined with analytical modeling and numerical studies to provide an understanding of the influence of reactions on film-cooling heat transfer.

The paper continues in Section 2 with an overview of the experiment, time scale considerations, and test hardware. Section 3 discusses the experimental tests and results. A numerical investigation of a film-cooled surface is presented in Section 4. Finally, Section 5 contains a data scaling method as well as the implications and utility of the investigation.

2 Transient Shock Tube Testing and Experimental Apparatus

To simulate conditions in which near wall reactions may occur a simplified experiment was devised and implemented. The purpose of this section is to discuss the experimental apparatus, relevant time scales, and data acquisition.

2.1 Shock Tube Experiment and Facility Overview. The shock tube provides a flexible, cost-effective facility for exploring a wide range of flow-chemistry interactions at the high temperatures associated with modern combustors. Sufficient time must be allowed for a quasi-steady flow to develop over the film-cooled flat plate so that relevant heat flux measurements may be obtained. However, the short run times are advantageous because the test specimen is exposed to high temperature flow for only a fraction of a second. The gas composition and test gas pressure are regulated to yield different stagnation temperatures and pressures. Similar transient techniques have been used by others to make nonreactive heat transfer measurements, [2].

The shock tube used in this experiment consists of a 7.3-m driven section and 8.4-m driver, both constructed from 30 cm diameter stainless steel pipe. Freestream pressure was measured by four pressure transducers located in the driven section, which also served to measure shock speed and infer total temperature. Argon was used as the test gas because of its high specific heat ratio, allowing for test temperatures up to 2800 K. Furthermore, it is inert and will not react with the seeded fuel. A choked nozzle located downstream of the test section served as a throttle to control freestream Mach number. Additional details concerning shock tubes and the MIT shock tube facility are found in reference [3].

Contributed by the International Gas Turbine Institute and presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Amsterdam, The Netherlands, June 3–6, 2002. Manuscript received by the IGTI, November 27, 2001. Paper No. 2002-GT-30182 Review Chair: E. Benvenuti.

Table 1 Test time scale summary (ms), ethylene fuel, $\Phi=1.0$, $P=6$ atm

T (K)	τ_{test} (ms)	u_{flow} (m/s)	τ_{flow} (ms, $x/D=10$)	$\text{C}_2\text{H}_4\tau_{\text{chem95}}$ (ms)	Da ($\tau_{\text{flow}}/\tau_{\text{chem95}}$)
1000	40	176	0.14	30	0.005
1900	22	243	0.11	0.08	1.4
2800	8	295	0.086	3e-3	28.7

2.2 Test Time Considerations. Calculations were performed to ensure that there would be ample time for the starting of flow through the test section and sufficient boundary layer development over the flat plate during the test. Table 1 summarizes the duration of the test, flow speed, and flow time over the range of temperatures studied. A thermal boundary layer analysis was performed which demonstrated that the initial ambient wall temperature would not deter ignition. Furthermore, it was shown that there is sufficient time for the boundary layer to reach a quasisteady state, [3,4].

A useful parameter to relate the flow and chemical time scales is the Damköhler number:

$$\text{Da} = \frac{\tau_{\text{flow}}}{\tau_{\text{chem}}} = \frac{L/u_{fs}}{\tau_{\text{chem}}} = \frac{10D/M\sqrt{\gamma RT_{\infty}}}{\tau_{\text{chem}}} \quad (1)$$

When this ratio is larger than unity, there is enough time for chemical reactions to take place during the traverse of the flow. Conversely, when this ratio is less than unity, the flow time is shorter than the required time to complete the reactions. In this study, τ_{flow} is defined as the time for the freestream to travel a distance of 10 cooling hole diameters, $L=10D$. This value was chosen for convenience and later it is shown that this also corresponds to the approximate location of the maximum augmentation in surface heat flux due to local reactions.

The primary objective of the facility was to probe a range of Damköhler numbers to represent conditions in current and future gas turbine technology as discussed in Part 1 of this study [1]. Ethylene, C_2H_4 , was chosen as the fuel because of its relatively short ignition time and ability to satisfy the shock matching condition [3].

Chemical ignition times for ethylene were estimated over the range of temperatures used in the experiment. Table 1 also presents the time to 95 percent completion, which was taken as ten times the chemical ignition time [1]. Examining the ratio of the flow to chemical time indicates that the chemical times are much shorter than the available test time, meaning that although the shock tunnel tests are short in duration, they are sufficiently long for full heat release of the fuel to take place.

Using the definitions in Eq. (1) and ethylene as the fuel, Damköhler numbers that range from 0–30 can be examined using the facility. In the experiment the Damköhler number was controlled through changing the freestream temperature. (In the numerator, T_{∞} , sets the freestream velocity since the test section Mach number is fixed at 0.3. In the denominator, for constant pressure and $\Phi=1.0$, the chemical time is only a function of temperature.)

Fuel content in the argon freestream was chosen to span a range of levels representative of energetic chemical emissions (portion of the composition that retains potential for further oxidation and heat release) due to combustor unmixedness, as was discussed in Part 1 of this study [1]. A 20,000 ppm CO equivalent case is representative of a modern combustor operating at $\Phi=0.5$ with 85 percent efficiency, or a future combustor operating at $\Phi=1.0$ with 100 percent efficiency. The 60,000 ppm CO equivalent case represents either a 1 sigma deviation in mixedness from a stoichiometric combustor or a 3 sigma deviation from a modern combustor with $\Phi=0.5$.

A more universal unit for fuel level, rather than equivalent CO concentration, is the ratio of the potential increase in enthalpy of the freestream due to the reaction relative to the freestream en-

thalpy. Since C_p for argon is a constant, this is expressed approximately in terms of total temperature, as shown in Eq. (2)

$$H^* = \frac{T_{\text{ad}} - T_{t\infty}}{T_{t\infty}} \quad (2)$$

The range of nondimensional fuel enthalpy studied in this investigation was 0.005–0.80. In the experiments the value of H^* increased with decreasing Damköhler number because the fuel mass fraction of the freestream was kept constant for each case. The results of the reacting flow tests will be shown in Section 3.4.

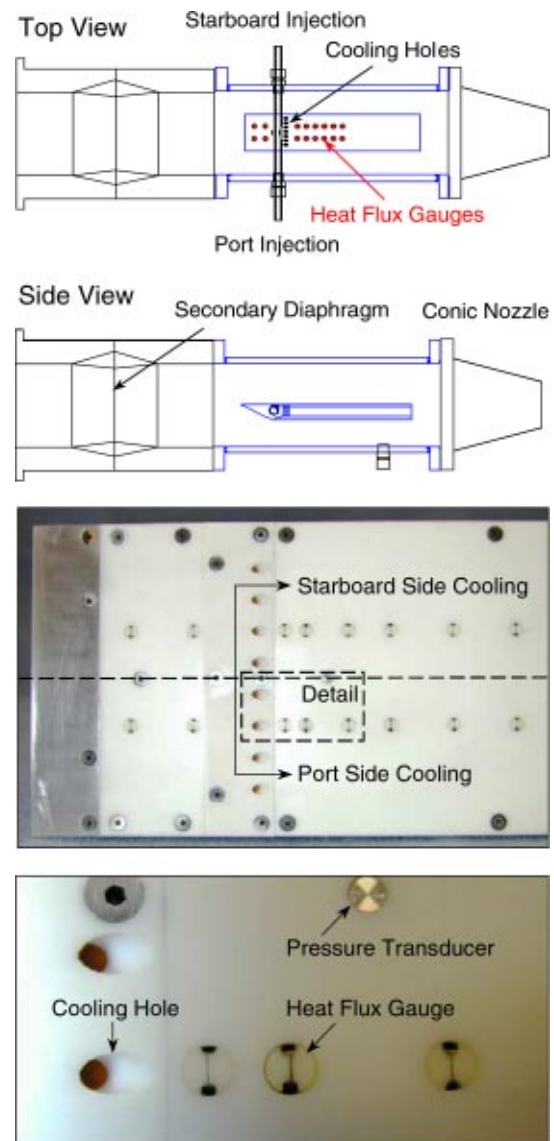
**Fig. 1 Test section and flat plate details**

Table 2 Summary of test cases

Test	P (atm)	T_∞ (K)	B (port)	B (starboard)	CO equivalent, ppm	H^*
1. Un-cooled flat plate tests						
1.1	5–7	1000–2800	N/A	N/A	N/A	N/A
2. Film-cooled, nonreactive tests						
2.1	6	1000–2800	1.0 (air)	0.0	N/A	N/A
2.2	6	1000–2800	0.0	1.0 (air)	N/A	N/A
2.3	6	1000–2800	0.5–2.0 (air)	0.5–2.0 (air)	N/A	N/A
2.4	6	1000–2800	0.5–2.0 (air)	0.5–2.0 (N ₂)	N/A	N/A
3. Reacting freestream flow tests						
3.1	6	1000–2800	0.5 (air)	0.5 (N ₂)	80,000–2,000	0.005–0.95
3.2	6	1000–2800	1.0 (air)	1.0 (N ₂)	60,000–5,000	0.01–0.80
3.3	6	1000–2800	2.0 (air)	2.0 (N ₂)	60,000–5,000	0.01–0.80
3.4	6	1500, 2800	1.0 (N ₂)	1.0 (air)	60,000, 20,000	0.1–0.80

2.3 Test Article Description, Data Acquisition and Processing. A schematic of the test section and detail of the flat plate test article are shown in Fig. 1. The flat plate used in this experiment is separated into two regions, such that two sets of 35 deg holes can be supplied with different film-cooling gases. One side of cooling holes was supplied with nitrogen and the opposite air. Four cooling holes were located on each side, with a diameter of 0.25 cm and a lateral spacing of 0.76 cm.

Thin-film heat flux gages were located upstream and downstream on each side of one of the film-cooling holes, with x/D locations ranging from -10 to 25 , as shown in Fig. 1. The gauges measure the surface wall temperature, which was integrated over time to give surface heat flux. The mixture of argon-ethylene provided a fuel rich freestream flow that reacted with the air side coolant, while simultaneously flowing passively over the inert nitrogen cooled side. This allowed for a same test, side-by-side assessment of the impact of local reaction. A hot-wire anemometer study was conducted to ensure that the supply plenum was providing uniform blowing through each set of cooling holes.

The test plate was constructed from aluminum and is 25.4 cm long by 7.62 cm wide. A 0.48-cm-thick sheet of MACOR ceramic was used as the plate surface due to its low thermal conductivity, k , (1.46 W/m K, versus 237 W/m K for pure aluminum) giving a high surface temperature rise (~ 50 – 80 K) during the 15 ms test duration. The heat flux gages were constructed from 0.05 cm MACOR cylinders. The thin resistive element was hand-painted on the top of each cylinder. Each gage, along with its signal amplifier, was calibrated in an oil bath after being mounted onto the test plate. This calibration produced a voltage versus temperature relationship, with the temperature known to better than 0.1 K.

The Reynolds number range at the cooling hole location varied from 1.3×10^5 to 2.4×10^5 depending on test condition. Care was taken when mounting the gages onto the test plate to ensure that the gage surface and the surface of the MACOR were to within 0.013 cm which satisfies the criteria for sandgrain aerodynamic smoothness. The transition distance to a turbulent boundary layer is on the order of 1 cm, putting it close to the first heat flux sensor [4,5]. A thin piece of tape located on the leading edge was used as a boundary layer trip.

The wall temperature and pressure data were acquired at 100 kHz. Wall temperature data from each individual gage were processed using the ACQ model. ACQ is an electrical circuit analogy model that was used in the integration of temperature versus time to heat flux. More information on the ACQ model can be found in reference [6].

3 Test Description and Experimental Results

The fundamental objective of the research program was to investigate the change in surface heat flux due to interaction of a fuel-rich freestream and a film-cooling layer. In order to proceed toward an assessment of local reactions and to build confidence in the transient heat flux data, the facility was first benchmarked against correlations and data available in the literature. Table 2

presents a summary of the types of tests completed to assess the performance of the facility, as well as those conducted with a reaction. Also shown in Table 2 are the range of blowing ratios studied in the film-cooled, nonreactive and reactive test series. The final column in Table 2 shows the fuel levels, in equivalent ppm CO and H^* .

A sample of the tunnel pressure traces, and a pressure trace just upstream of the test plate are shown in Fig. 2 for a freestream temperature of 2000 K. From the pressure trace, the location of the shock as it moves through the tunnel can be seen on both the incident and reflected passes. The vertical lines indicate the region taken as quasi-steady, based on the uniformity of the pressure traces. For this nonreactive test, the region of relevance is approximately 9 ms in duration at a pressure of 6 atm.

The second plot shows the wall temperature response of the two heat flux gages located transverse to each other at 3.2 in., $x/D = 32$, downstream of the leading edge. The plot indicates a rapid increase in temperature over the start-up time, and then a gradual increase in temperature during the actual test window. This wall temperature versus time trace was then integrated to give the wall heat flux, which is shown in the third plot to be around 5×10^4 W/m² during the test window.

3.1 Transient Flat Plate Heat Transfer Physics. This section explains why the temperature and heat flux versus time traces look the way they do in Fig. 2. At the outset of the test, a shear flow develops over the flat plate, and since no initial boundary layer is present, the high heat load to the plate causes the wall temperature to rise rapidly. Shortly thereafter a boundary layer develops, which buffers the heat load to the wall. Depending on the magnitude of the initial heat flux and the heat flux over the quasi-steady region of the test, the wall temperature may continue to rise, remain level, or even decrease depending on the balance of convective and conductive heat transfer over and into the wall, respectively. The temperature distribution within a semi-infinite solid for a step change in surface heat flux is given by

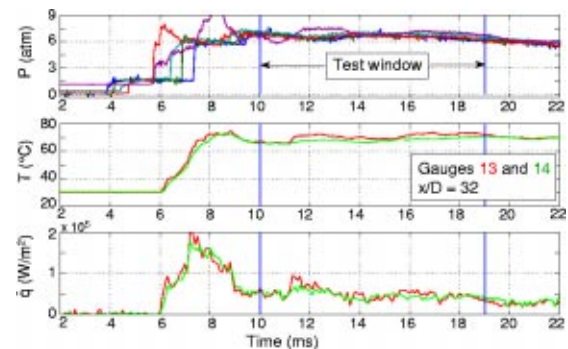


Fig. 2 Sample pressure, wall temperature, and wall heat flux traces, $T_\infty = 2000$ K

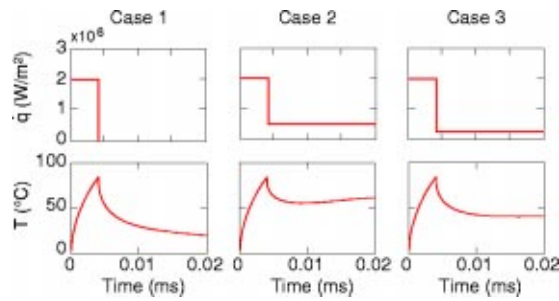


Fig. 3 Wall temperature distributions ($\zeta=0.0$) for three wall heat flux distributions

$$\Delta T = \frac{2\dot{q}}{k} \left[\left(\frac{\kappa t}{\pi} \right)^{0.5} \exp\left(\frac{-s^2}{4t\kappa}\right) - \frac{s}{2} \operatorname{erfc}\left(\frac{s}{2\sqrt{\kappa t}}\right) \right] \quad (3)$$

Figure 3 provides an example of an initial heat load of $2 \times 10^6 \text{ W/m}^2$ to model the shear start-up flow and then considers three different levels of heat load for a fully developed turbulent boundary layer at the wall ($\zeta=0.0$).

In case 1, the initial heat load is applied for 4 ms and removed with no subsequent heat flux. The temperature response is an initial spike and then decay. In case 2, the steady-state heat load is just large enough to make the wall temperature continue to climb after the initial shear layer spike. In case 3, it is seen that if the proper steady-state heat load is applied, the wall temperature will spike and then remain effectively constant over the duration of the test window. Comparing case 3 with Fig. 2, the physical interpretation of the relationship between the temperature response and integrated heat flux is elucidated. To appropriately include these effects, the entire temperature versus time history is used to integrate the wall temperature during the quasi-steady window.

3.2 Error Analysis. This section discusses the errors and uncertainties associated with making transient heat flux measurements. Two types of errors are considered: 1) absolute error in the calculated quantities, and 2) differential error in evaluating relative changes between the air and N_2 sides of the plate. In the experiment, the wall temperature is known to better than 2 percent, and hence the integrated heat flux is known to better than 2 percent. The freestream temperature is calculated from the shock speed. Although the shock speed is known to better than 0.5 percent imperfect reflections, viscous effects, and other secondary flow features create uncertainties in total temperature on the order of approximately 10 percent. The horizontal bars in Fig. 7 relate an offset bias in temperature to Damköhler number. Further, although the heat flux and wall temperature are well known, the absolute heat transfer coefficient is only known to within 10 percent, which is consistent with the run-to-run repeatability of the facility. However, the differential experiment allows for an assessment of the change in the heat transfer coefficient (between the two sides) to better than 5 percent.

3.3 Un-Cooled Flat Plate Heat Transfer Results. In order to benchmark the experiment, a series of uncooled flat plate tests were conducted, as shown in Table 2. The purpose of these tests was to ascertain how well the measured heat flux agreed with analytical correlations for turbulent flat plate heat transfer. The turbulent flat plate heat flux correlation given in reference [4] is

$$\dot{q} = h(T_\infty - T_0) = \rho C_p \text{St} M \sqrt{\gamma R T_\infty} (T_\infty - T_0) \quad (4)$$

with the Stanton number distribution given by

$$\text{St} = \frac{0.0287 \text{Re}_x^{-0.2}}{0.169 \text{Re}_x^{-0.1} (13.2 \text{Pr} - 10.16) + 0.9} \quad (5)$$

Figure 4 is a summary of a representative set of cases over the

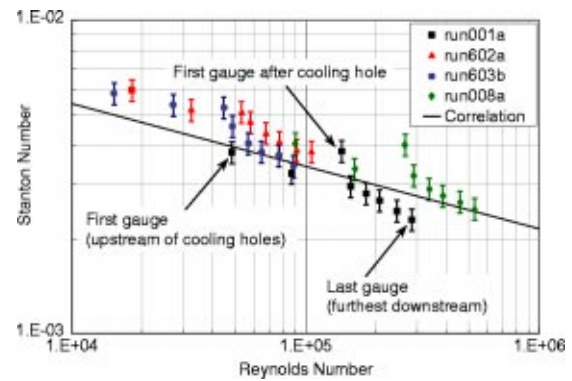


Fig. 4 Flat plate heat transfer results

temperature range investigated, with the symbols representing the average heat flux taken at each gauge location over the quasi-steady portion of the test. The solid line shows the predicted turbulent heat flux on a flat plate. The plot indicates that the overall heat flux levels agree to within 10–15 percent of the correlation. The vertical bars shown at each heat flux gage location represent the peak fluctuations within the steady-state window, as shown in the bottom plot of Fig. 2.

These fluctuations are on the order of 10 percent of the time-averaged value. Side-to-side agreement between lateral heat flux gages at the same x/D location was on the order of 10 percent. For each test, the points which show the greatest deviation are located just aft of the cooling holes. During these tests a piece of tape was placed over the cooling holes to prevent back-flow and reduce flow disturbances due to the cooling holes. This may explain why these points consistently show deviation. Nonetheless, the data show good agreement both in trend and magnitude with Eqs. (4) and (5).

3.4 Film-Cooling Tests. The next set of validation tests was designed to assess the performance of transient film-cooled tests. Film-cooled flat plate experiments were conducted to ensure side-to-side agreement of air and nitrogen over a range of blowing parameters from 0.5–2.0. A comparison of experimental film-cooled results to data in the literature was performed [7]. An example is shown in Fig. 5 for one side with no blowing, and the other having $B=1.0$. The side with a blowing parameter of 1.0 shows a 10–35 percent reduction in surface heat flux, depending on downstream distance from the cooling holes. Also note the two side-by-side gages upstream of the cooling holes are consistent with each other. The differential experiment is valuable for this type of comparison. Instead of conducting two experiments at

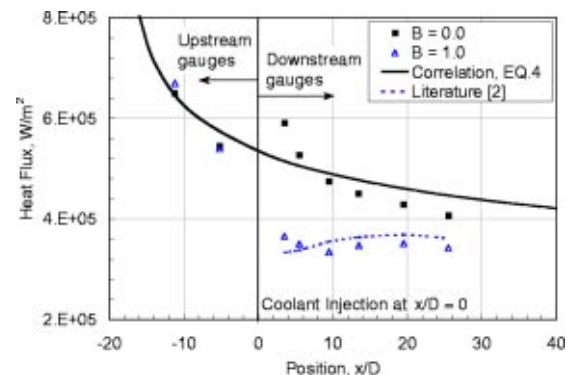


Fig. 5 Sample of film-cooling test result

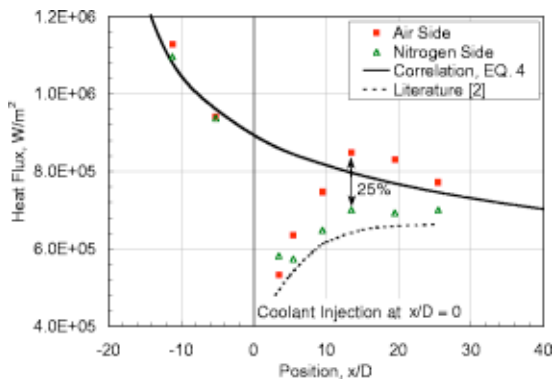


Fig. 6 Augmented heat flux due to local reaction, $B=1.0$, $H^*=0.18$, $Da=20$

identical freestream pressure and temperature between runs, a direct assessment of the influence of different blowing parameters is available in a single test.

In order to assess the cooling performance, literature data for film-cooling effectiveness were used to provide an approximate comparison. To arrive at an estimate, adiabatic film-cooling effectiveness versus x/D was used. The adiabatic film-cooling effectiveness is defined as

$$\eta_{ad} = \frac{T_{\infty} - T_f}{T_{\infty} - T_c} \quad (6)$$

The dashed line in Fig. 5 represents the expected improvement from film-cooling effectiveness as applied to the data from the side with no cooling. This is done by multiplying the uncooled side by $(1 - \eta_{ad})$ to arrive at an estimate for the predicted reduction in heat flux due to film-cooling. Over the range of blowing parameters investigated, the performance of the film-cooled side agreed with the literature effectiveness data to about 5–30 percent. It should be noted that the data taken from reference [7] is not an exact match. The literature data is for air injection into an air freestream and the Reynolds numbers are not identical. However, the agreement is equally good when compared to the nonreacting numerical simulations of the experiment discussed in Section 4.1.

3.5 Reactive Flow Tests. This section is concerned with the primary goal of the experiment: measuring the change in surface heat flux due to local reaction over a range of Damköhler numbers. A series of tests using ethylene fuel, $H^*=0.005-0.8$, seeded in the argon freestream were conducted to assess the change in surface heat flux due to local reaction. These fuel levels were chosen as a representative range of the levels of chemical emissions due to combustor unmixedness [1]. Blowing ratios of 0.5–2.0 and freestream temperatures of 1000–2800K were examined, corresponding to a Damköhler number range of 0–30. The complete series of reactive flow tests conducted is summarized in Table 2. Figure 6 shows a sample of the influence of local reaction on measured surface heat flux. Air was injected on one side and nitrogen on the other side, both at a blowing ratio of 1.0. The freestream temperature was 2200K with a nondimensional fuel enthalpy of $H^*=0.18$ and $Da=20$. The differential experiment allows for a back-to-back comparison of the heat fluxes associated with the reactive side (air injection) with the inert side (N_2 injection).

Examining Fig. 6, differences of up to 25 percent exist between the reactive and nonreactive sides downstream of the film-cooling holes, whereas the agreement between measured heat flux on the upstream gages is within several percent. The plot indicates the influence of the reaction extends downstream to the last gauge location at $x/D=25$. The solid line shows the empirical turbulent

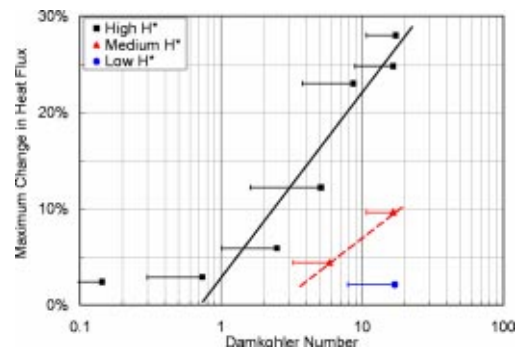


Fig. 7 Maximum change, $B=1.0$, $H^*=0.005-0.8$

boundary layer correlation (Eq. (4)). The agreement of the nitrogen (inert) cooled side to the literature predicted decrease in surface heat flux, which is shown by the dash lined and computed as described in Section 3.3, is within 15 percent. To ensure that this behavior is attributable to local reactions, the gas injection sides were switched and the experiment repeated with similar results.

3.6 Summary of Reacting Flow Results. Tests of the type shown in Fig. 6 were conducted for momentum blowing ratios of 0.5, 1.0 and 2.0 over a range of Damköhler numbers at three different fuel enthalpy levels—high: $H^*=0.18-0.8$, medium: $H^*=0.06-0.24$, and low: $H^*=0.005-0.03$. For each fuel level tested, the maximum difference in side-to-side heat flux was located at an x/D of 10–15. The change in side-to-side heat flux increased with Damköhler number as shown in Fig. 7 for a blowing ratio of 1.0 for all fuel enthalpies.

At the highest Damköhler numbers tested, the cases with high H^* had the most heat flux augmentation of around 30 percent. Similarly, the medium H^* cases also exhibited the maximum change in side-to-side heat flux at the highest Damköhler numbers, however these levels were on the order of 10–15 percent change. No measurable change in side-to-side heat flux above the experimental uncertainty limit was observed in the low H^* cases. In summary, the largest change in side-to-side heat flux for any fuel level always occurred at the highest Damköhler number.

Figure 8 compares the differences in the heat flux signatures for an attached jet ($B=0.5$) and lifted jet ($B=2.0$), at a Damköhler number of 24. Both cases exhibit peak heat flux at an x/D of around 10–15, but with the lifted side peaking somewhat further upstream. The augmented heat flux for the attached case decreases slowly after peaking, whereas the lifted jet tends to decrease quicker after peaking.

This is qualitatively understood because in the attached case more time is required for the heat within the film layer to diffuse and mix to the wall leading to increased heat flux downstream.

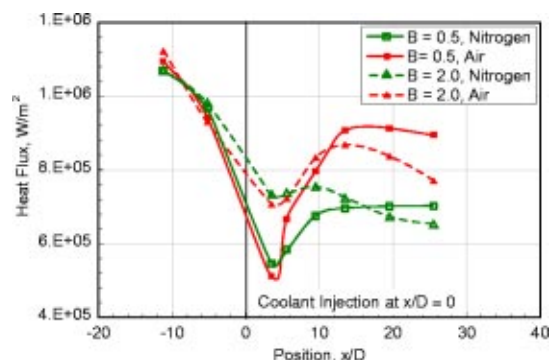


Fig. 8 Comparison of $B=0.5$ versus 2.0, $H^*=0.18$, $Da=24$

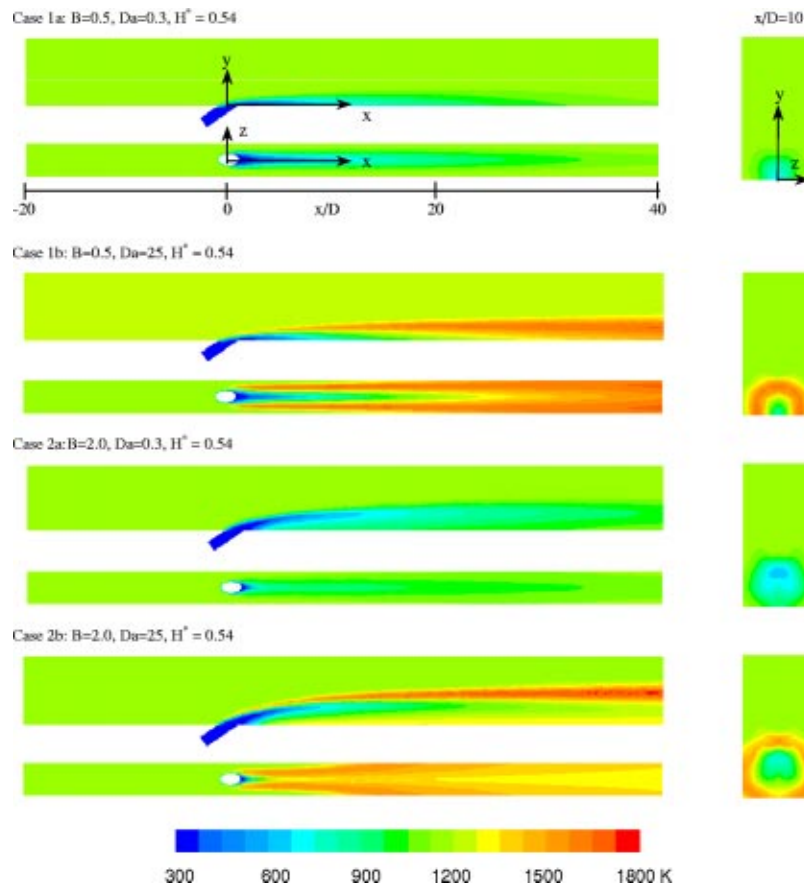


Fig. 9 Numerical study results for adiabatic wall cases

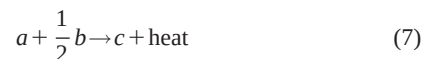
However, in the lifted case, hot gases are entrained towards the wall as a consequence of recirculation behind the jet. This allows for increased potential heat release directly behind the film cooling hole. This mixing with the freestream is also responsible for reducing the temperature of the hot gases, leading to a lower wall heat flux for the same freestream fuel level. This behavior is further discussed in Section 4 where numerical simulations are used to study differences in locally reacting attached and lifted jets.

4 Numerical Simulations

The objective of the numerical simulations was to provide a means to quickly scope the general behavior of a film-cooled reacting flow field. This included assessing differences in the behavior of attached and lifted jets for a range of Damköhler numbers, blowing parameters, freestream fuel levels, identifying the location of primary heat release on and off the surface, and determining how film effectiveness is altered in the presence of local reactions. The studies were designed to provide a basic understanding of the phenomena associated with locally reacting film cooling flows. Furthermore, the studies were designed to provide an understanding of how much of the change in heat flux due to local reactions was attributable to driving temperature and how much was due to a change in heat transfer coefficient.

Fluent was used to model the flow field. A tetrahedral grid was created for a single 35 deg film-cooling hole. The grid domain extended 20 and 40D upstream and downstream of the cooling hole, respectively. The simulation was performed using a second-order discretization scheme. Turbulence away from the walls was modeled using the standard $k-\varepsilon$ equations. Near the wall, Fluent's standard wall function model was implemented which encompasses law of the wall scaling and near wall turbulence modeling [8,9]. Convergence was determined based on overall mass and

energy imbalance within the domain of less than 0.01 percent. To model the chemistry, a simple 1-step irreversible reaction model was employed as summarized in the following:



$$\tau_{\text{chem}} \approx \frac{[\text{fuel}]_0}{A e^{-E_a/RT_0} [\text{fuel}]^a [\text{O}_2]^b} \quad (8)$$

The freestream flow was argon seeded with hydrogen, a , and the cooling flow was air, b . The Da number sweep was performed through varying the pre-exponential factor, A .

4.1 Numerical Results. A numerical study was completed over a range of blowing parameters and Damköhler numbers. Figure 9 presents a comparison between attached ($B=0.5$) and lifted ($B=2.0$) jets in nonreactive ($Da=0.3$) and highly reactive ($Da=25$) cases for $H^*=0.54$ with an adiabatic wall. In each set, the upper contour plot shows a cross section through $z/D=0$, and the lower plot is a cross section at the wall, $y/D=0$. In each case, the freestream temperature was 1200 K, the coolant air temperature was 300 K, and the freestream velocity was 100 m/s. Cases 1a and 2a show non-reacting film cooling flows for jets at blowing parameters of 0.5 and 2.0, respectively. The results of these numerical simulations were compared with literature predictions for film effectiveness [7] and shown to agree within 20 percent.

Cases 1b and 2b demonstrate the result of increasing the Damköhler number. The adiabatic flame temperature, T_{ad} , based on $\Phi=1.0$, is 1850 K. The maximum temperature in the flow field downstream of the film cooling holes reaches 1722 K. Attached jet temperature contours show the region of increased temperature is concentrated near the wall and located at the edges of the film

cooling layer. The region of augmented heat flux then propagates to the centerline at around $x/D=10$ and continually increases downstream. The lifted jet shows two distinct regions of reaction, on the top and bottom of the jet structure. The reactions take place circumferentially around the cooling jet as can be seen in the $x/D=10$ cross-sectional view of Fig. 9. The peak temperature in each of these zones is 1683 K. For the lifted jet, the increased temperature region near the wall is concentrated between $3 < x/D < 15$, and then decreases with downstream distance. Qualitatively, this behavior is in agreement with the experimental data of Fig. 6.

Also shown in Fig. 9 are cross sections taken at an $x/D=10$. These plots indicate that for the attached jet the peak wall temperature does not occur on the centerline behind the hole but rather around $z/D = \pm 1.0$. In contrast, the lifted jet always exhibits a peak wall temperature increase on the centerline, $z/D=0$. This can be understood because the location of the reactive interface between the freestream and the coolant is different for attached and lifted jets. In the attached case this interface meets the wall off-centerline, whereas for lifted jets this interface is closest to the wall at $z/D=0$. The difference between the wall temperature at $z/D=0$ and $z/D=\pm 1.0$ for the attached cases, at $x/D=10$ is about 5 percent.

One goal of the numerical studies was to ascertain how much of the augmentation in heat flux due to local reactions was attributable to a change in driving temperature and how much was due to a change in heat transfer coefficient, $h = q_w / (T_\infty - T_w)$. To investigate the changes in h the wall temperature was varied from 300 K (coolant temperature) to 600 K (freestream temperature) with and without reaction. As the wall temperature was increased the percent change in spatially-averaged heat transfer coefficient, $(h_{\text{hot}} - h_{\text{cold}}) / h_{\text{cold}}$, varied from -5 to $+50$ percent. Results for a reacting boundary layer, where peak temperatures occur approximately midway into the shear layer, suggest that a factor of two increase in the heat transfer coefficient is possible due to profile effects [10].

The numerical studies were also used to determine the change in film effectiveness due to local reaction. The film effectiveness degrades significantly in the presence of local reactions, and can become negative. This will be shown in more detail in Fig. 12 in the subsequent section. Finally, the wall temperature did not exhibit a significant influence on the magnitude or location of the reactions.

5 Data Summary, Data Scaling, and Applications

The goals of this section are twofold. First, to describe non-dimensional scaling parameters that enable the experimental and numerical data to be used to predict heat loads for other freestream fluid and chemical conditions. Second, to demonstrate through two case study examples the potential impact of secondary reactions for representative gas turbine engine scenarios.

5.1 Experimental Data Scaling. In order to apply the experimental results, the data set must be scaled so that it can be referenced to other freestream conditions. This is accomplished using the following scaling:

$$Q_s = \frac{q_{\text{hot}} - q_{\text{cold}}}{q_{\text{max}} - q_{\text{cold}}} = \frac{q_{\text{hot}} - q_{\text{cold}}}{q(T_{\text{ad}}) - q_{\text{cold}}} \quad (9)$$

The maximum possible heat flux, q_{max} , is that which would be achieved if the local reaction proceeded to the adiabatic flame temperature, T_{ad} . This is evaluated by setting $T_\infty = T_{\text{ad}}$ in Eq. (4). The scaled heat flux, Q_s , represents how much surface heat flux augmentation was achieved for any given test compared to how much augmentation could possibly have been achieved. The q_{hot} and q_{cold} heat fluxes were directly evaluated from the measured wall temperature versus time histories in the experiment as was shown in Fig. 2. For each test, the adiabatic flame temperature was calculated using freestream temperature and pressure and as-

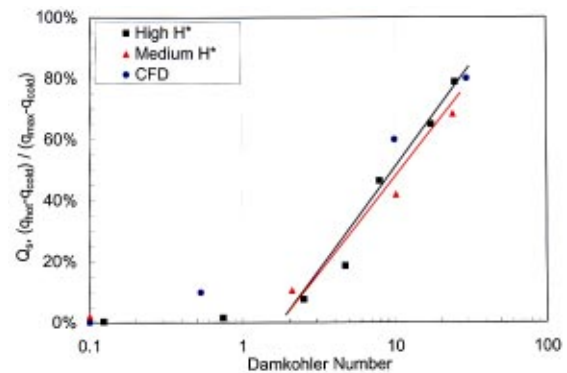


Fig. 10 Scaled data set for $B=1.0$, $H^*=0.05-0.8$

suming that somewhere in the flow field the local equivalence ratio was unity. Thus if $q_{\text{hot}} = q_{\text{cold}}$, 0 percent of the potential augmentation was achieved, and if $q_{\text{hot}} = q_{\text{max}}$, 100 percent of the possible augmentation was achieved.

The scaled experimental data are shown in Fig. 10 for $B=1.0$. At the highest Damköhler number that can be captured in the experiment, the scaled heat flux is on the order of 70–80 percent. Straight lines have been fitted to the data to show the general trend, although it is expected to asymptote to 100 percent as the Damköhler number is further increased.

At any Damköhler number, the near wall reactions tended to achieve the same level of scaled heat flux regardless of the non-dimensional heat release potential of the fuel. This was corroborated by a series of numerical experiments as described in Section 4.1. Studies were conducted at $B=1.0$ and constant Damköhler number, while H^* was varied from 0.15–0.54. For each value of H^* , the scaled heat flux remained constant. Furthermore, the value of Q_s was shown to increase with increasing Damköhler number to a value of 85 percent at $Da=25$, consistent with the experimental trends. The results of the numerical study are also shown in Fig. 10. The fuel level, or H^* , is important because it indicates the severity of the heat flux augmentation in terms of potential change in ΔT_t . (Note that the lowest fuel levels, $H^* < 0.03$, are not shown in Fig. 10 because the change in side-to-side Q_s was smaller than the experimental uncertainty).

The effect of blowing ratio on the scaled experimental heat transfer data is shown in Fig. 11 for high H^* (0.18–0.8). Each data point shown in the figure represents the maximum side-to-side change that was shown in Fig. 6 to occur at an x/D of around 10–15. As was described in Sections 3.6 and 4.1, the greatest heat flux augmentation for a given Damköhler number and fuel level occurred with the attached $B=0.5$ case. The scaled heat flux de-

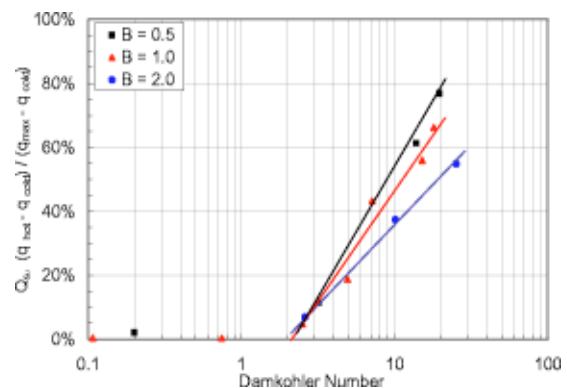


Fig. 11 Scaled data set for $H^*=0.18-0.8$

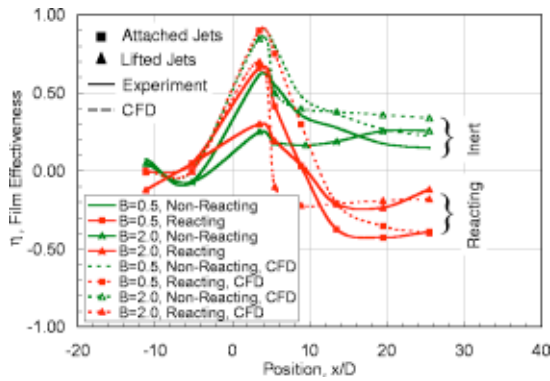


Fig. 12 Decay in film effectiveness due to local reactions, $H^* = 0.54$, $Da=25$

creased as the blowing ratio was increased. The side-to-side variation in heat flux due to local reaction becomes important at Damköhler numbers of around 2–5 for each of the blowing ratios examined.

The experimental data can also be assessed in terms of film effectiveness. The wall heat flux is defined as the convective heat transfer coefficient times the difference between some driving temperature (which can range between T_∞ and T_f) and the wall temperature. For a measured change in surface heat flux, $q_{hot} - q_{cold}$, it is possible to estimate the change in driving temperature if the behavior of the heat transfer coefficient with temperature and reaction is known. The numerical studies presented in Section 4.1 indicated that the change in h due to local reactions is between -5 and $+50$ percent depending on wall temperature and x/D location. Since these changes in h typically contribute less to the change in surface heat flux than the increase in driving temperature, the heat transfer coefficient is taken to be constant in the forthcoming analyses (this is particularly true for the experimental wall temperatures). By solving Eqs. (10a) and (10b) for the driving temperatures and knowing the wall and coolant temperature, the film effectiveness can be inferred for the hot and cold cases.

$$q_{cold} = h(T_{dc})[T_{dc} - T_{wc}] \quad (10a)$$

$$q_{hot} = h(T_{dh})[T_{dh} - T_{wh}] \quad (10b)$$

$$q_{max} = h(T_{ad})[T_{ad} - T_w] \quad (10c)$$

Figure 12 shows agreement between the film effectiveness derived from the experiment (solid lines) and that from numerical simulations (dashed lines) for attached and lifted jets at low and high Damköhler numbers. Attached jets are shown with square symbols, and lifted jets with triangles. Note that the film effectiveness may become negative in the presence of local reactions. This is understood by examining Equation 6 and noting that the film, or driving temperature, may exceed freestream temperatures when local reactions are present, causing the effectiveness to drop below zero.

For the attached jet, the decay in effectiveness monotonically declines with increasing x/D , with the reacting flow case decreasing more rapidly to -0.4 at an $x/D = 25$. The effectiveness crosses zero at an x/D of around 15. For the lifted jet, the effectiveness peaks just downstream of the cooling hole and then rapidly decreases before increasing as the jet begins to reattach downstream. The effectiveness crosses zero at an x/D of approximately 5. The same phenomenon is seen to occur in the reacting case, but once again with the effect being intensified with effectiveness down to -0.5 at an $x/D = 15$. Both the experiment and the numerical results show an increase in effectiveness downstream of $x/D = 15$ for the lifted jet, which is a result of reattachment of the film

cooling layer further downstream. For the attached jet the experimentally inferred effectiveness and the numerical prediction do not exhibit this behavior.

5.2 Application of Results to Typical Film-Cooling Scenarios. Here we consider the application of the preceding data, as well as the analyses of Part 1 of this study [1] to two film-cooled surface design applications. The conditions used in these examples are representative of advanced technology gas turbine environments.

Case 1: Airfoil with Film Cooling. To determine the effect local reactions have on the heat flux to a blade surface, we consider a film-cooled turbine blade operating at $B=2.0$ with a film effectiveness of 0.2. Velocity over the blade is 100 m/s, freestream temperature and pressure are 2000 K and 25 atm., respectively, the coolant temperature is 950 K, and the wall temperature is 1200 K. We assume that fuel rich streaks with an energy of 60,000 ppm CO equivalent have convected from the combustor to the blade surface. Such a case is analogous to a fuel rich streak with an equivalence ratio of 1.2 exiting the combustor.

We begin by calculating a representative Damköhler number for this flow scenario using Eq. (1). Taking a typical spacing of 1 cm between film cooling holes on the blade, the flow time over this distance is 0.0001 s. An adiabatic, premixed calculation (as discussed in Part 1 of this study) is then performed to arrive at the chemical time and maximum total temperature change. The τ_{chem} and ΔT_t for the freestream temperature, pressure, and fuel content are 0.00004 seconds and 350 K, respectively [1]. This gives a Damköhler number of approximately 3 and non-dimensional heat release potential of $H^*=0.18$. From Fig. 11, we see that the scaled heat flux ratio, Q_s , is about 0.2. Given $\eta=0.2$ for this scenario, we can calculate T_f from Equation 6, which will be used as the representative driving temperature, T_{dc} ($T_f \sim T_{dc} = 1630$ K).

Next we calculate q_{cold} and q_{max} for the given conditions. The heat transfer coefficient is calculated using the flat plate correlation (Eq. (4)) and is about 1700 W/m² K. Again for simplicity we assume that the change in heat transfer coefficient with driving temperature is negligible in this analysis. We could instead assume a change in heat transfer coefficient as indicated by the numerical studies. Note that any assumed increase in the heat transfer coefficient with reaction would further magnify the estimated augmentation in heat flux due to reaction. Note that in the calculation for q_{max} , the film effectiveness is allowed to go to zero to arrive at a maximum wall heat flux. (Numerical analyses of this type of cooling jet in these freestream conditions show an expected decrease in film effectiveness from 0.2 to around zero.)

$$q_{cold} = h(T_{dc})[T_{dc} - T_{wc}] = 1700[1630 - 1200] = 7.3 \times 10^5 \text{ W/m}^2$$

$$q_{max} = h(T_{ad})[T_{ad} - T_w] = 1700[2350 - 1200] = 2.0 \times 10^6 \text{ W/m}^2$$

Solving Eq. (9) gives $q_{hot} = 9.9 \times 10^5 \text{ W/m}^2$. This simplified analysis indicates that a designer would expect 35 percent more heat flux to the blade surface due to the presence of local reactions.

Case 2: Combustor Film-Cooled Surfaces. In this example we will proceed in the same manner as Case 1, but with different operating conditions to illustrate a larger potential change in q_{hot}/q_{cold} . We consider a film cooled combustor liner operating at a $B=1.0$, with a film effectiveness of 0.7. Velocity in this region is 10 m/s, freestream temperature and pressure are 1800 K and 20 atm., respectively, the coolant temperature is 950 K, and the wall temperature is 1200 K. Again we assume that a freestream with an energy equivalent of 60,000 ppm CO is present. In this case the spacing is 2 cm between film cooling holes, and the flow time over this distance is 0.002 s.

For these conditions $Da=20$, $\Delta T_t=300$ K, and $H^*=0.17$. From Fig. 11, the scaled heat flux ratio is about 0.75. Given $\eta=0.7$ for this scenario we can calculate T_f from Eq. (5) ($T_f \sim T_{dc} = 1290$ K). Again, we calculate q_{cold} ($1.3 \times 10^5 \text{ W/m}^2$) and

$q_{\max}(1.4 \times 10^6 \text{ W/m}^2)$ for the given conditions. Solving Eq. (9) gives $q_{\text{hot}} = 1.1 \times 10^6 \text{ W/m}^2$. In this case, a factor of 10 increase heat flux over the non-reactive case would be expected due to the presence of local reactions.

In each of these two cases, it is important to note the sensitivity of q_{hot} and q_{cold} to T_w , T_f , and T_c . As the film temperature approaches the wall temperature, q_{cold} decreases, making the denominator in the ratio of $q_{\text{hot}}/q_{\text{cold}}$ small. So even in the cases where the Damköhler number is of the order 1 and the scaled heat flux factor around 0.1, the ratio of $q_{\text{hot}}/q_{\text{cold}}$ can become large. However, the additional heat load to the film-cooled surface may not be significant in terms of ΔT_t . Therefore scaling with $q_{\text{hot}}/q_{\text{cold}}$ can be misleading and hence the scaling shown in Eq. (9) (based on percent of maximum achievable heat flux) is suggested when making comparisons between nonreactive and reactive scenarios. To estimate the severity of the impact to the film-cooled surface, B , Da , Q_s , and H^* must be known.

Summary and Conclusions

As the equivalence ratios within modern combustors increase, the potential for reactive species to interact at film-cooled locations must be taken into account during design of gas turbine engines. This paper presented a method for exploring the impacts that secondary reactions have on a film-cooled flat plate and showed how this data can be scaled and applied to simplified design scenarios. In summary:

- A simplified method utilizing experimental data, reduced-order analyses, and numerical results can be used to consider the influence of local reactions. Under certain operating conditions, augmentations in surface heat flux can be produced due to chemical reactions in the film-cooling layer.
- The impact of secondary reactions on a film-cooled surface is primarily a function of 4 non-dimensional quantities: The Damköhler number (Da), heat release potential (H^*), scaled heat flux (Q_s), and blowing ratio (B).
- The scaled data indicate that the scaled heat flux, Q_s , increases with increasing Damköhler number to a value of around 80 percent at $Da=30$ for $B=1.0$.
- Numerical reacting flow studies were performed demonstrating that for attached jets the location of greatest surface heat flux augmentation occurs for $x/D > 10$, at $z/D = \pm 1$. For lifted jets, increased heat load is concentrated between $3 < x/D < 15$ on the centerline. The film effectiveness degrades significantly in the presence of local reactions.

Future computational studies will focus on multiple row and staggered hole cooling geometries to examine how well single hole results extend to such configurations. Additionally, step geometries will be explored computationally and experimentally. Local reactions near step and recirculation zones, where τ_{flow} is very large, may exhibit substantially different behavior as compared to the single film cooling hole studies presented herein.

Acknowledgments

This material is based upon work under a National Science Foundation Graduate Fellowship. This effort was sponsored by the generous support of Pratt & Whitney. The authors would like to thank Drs. Frederick Soechting, Charles Graves, Saadat Syed, Thomas Auxier, William Kvasnak, and Atul Kohli of Pratt & Whitney, Dr. Catalin Fotache of the United Technologies Research Center, and Professor Edward Greitzer of MIT for their guidance and critique of this research. We also thank Mr. Krzysztof Fidkowski, MIT, for setting up and executing many of the numerical studies.

Nomenclature

A	= pre-exponential factor
B	= blowing parameter, $\rho_c u_c / \rho_\infty u_\infty$
C_p	= constant pressure specific heat (J/kg K)
Da	= Damköhler no. $\tau_{\text{flow}} / \tau_{\text{chem}}$
D	= cooling holes diameter (m)
E_a	= activation energy
h	= heat transfer coefficient (W/m ² K)
H^*	= heat release potential $(T_{\text{ad}} - T_{t\infty}) / T_{t\infty}$
k	= thermal conductivity (W/m K)
L	= length scale (m)
M	= Mach no. u_∞ / a_∞
P	= static pressure (N/m ²)
Pr	= Prandtl no. $C_p \mu / k$
$\dot{q}$	= heat flux (W/m ²)
Re	= Reynolds no. $\rho u_\infty x / \mu$
Q_s	= scaled heat flux ratio $(q_{\text{hot}} - q_{\text{cold}}) / (q_{\text{max}} - q_{\text{cold}})$
St	= Stanton no. $h / u_\infty \rho C_p$
T	= temperature (K)
x	= streamwise distance (m)
y	= vertical (normal to surface) distance (m)
z	= lateral distance (m)
Φ	= equivalence ratio
η	= film-cooling effectiveness $(T_\infty - T_f) / (T_\infty - T_c)$
κ	= thermal diffusivity (m ² /s)
μ	= viscosity (N s/m ²)
ρ	= density (kg/m ³)
τ	= characteristic time (s)
ζ	= penetration depth (m)

Subscripts

ad	= adiabatic
c	= coolant
chem	= chemical time to 95 percent completion
cold	= nonreactive heat flux
d	= driving
f	= film
hot	= reactive heat flux
max	= maximum heat flux
t	= total or stagnation quantities
test	= test time
w	= wall
∞	= freestream

References

- [1] Lukachko, S. P., et al., "Turbine Durability Impacts of High Fuel-Air Ratio Combustors, Part 1: Potential For Intra-Turbine Oxidation of Partially-Reacted Fuel," ASME GT-2002-30077.
- [2] Kercher, D. M., "Short Duration Heat Transfer Studies at High Free-Stream Temperatures," ASME 82-GT-129.
- [3] Kirk, D. R., 1999, "Aeroacoustic Measurement and Analysis of Transient Hot Supersonic Nozzle Flows," Master's thesis, Massachusetts Institute of Technology.
- [4] Keyes, W. M., and Crawford, M. E., 1980, *Convective Heat and Mass Transfer*, McGraw-Hill Book Company, New York, NY.
- [5] White, F. M., 1991, *Viscous Fluid Flow*, McGraw-Hill Book Company, Boston, MA.
- [6] Schultz, D. L., and Jones, T. V., 1973, "Heat-Transfer Measurements in Short-Duration Hypersonic Facilities," AGARD No. 165.
- [7] *Handbook of Heat Transfer Applications*, 1985, 2nd Ed., McGraw-Hill, New York, NY.
- [8] Walters, D. K., and Leylek, J. H., 2000, "A Detailed Analysis of Film-Cooling Physics: Part 1—Streamwise Injection With Cylindrical Holes," ASME J. Turbomach., **122**, Jan.
- [9] Walters, D. K., and Leylek, J. H., 1997, "A Systematic Computational Methodology Applied to a Three-Dimensional, Film-Cooling Flowfield," ASME J. Turbomach., **119**, Oct., pp. 777–785.
- [10] Eammons, H. W., 1956, "The Film Combustion of Liquid Fuel," Z. Angew. Math. Mech., **36**:1–2, pp. 60–71.

Assessment of Current Turbine Engine High Cycle Fatigue Test Methods

K. L. Nichol

Experimental Design & Analysis Solutions, Inc.,
318 Shadow Creek Drive,
Brentwood, TN 37027

Testing of turbine engine components to verify structural durability goals is performed in a variety of ways and at a variety of facilities. Further, methods used differ between engine companies and even among engine programs within the same company. A recent emphasis on reducing high cycle fatigue (HCF) failures of engine components has made it obvious that improvements and standardization of HCF test methods are needed. This paper presents findings from an assessment of current HCF test approaches and identifies weaknesses that form the basis for recommending improvements. Weaknesses identified included data processing, mode identification, and selection of instrumentation locations. [DOI: 10.1115/1.1423913]

Introduction

In cooperation with a national program aimed at reducing occurrences of high cycle fatigue (HCF) in fielded engine systems ([1,2]), an HCF test protocol working group from across the engine community was formed. This group was tasked with reviewing current HCF test and evaluation (T&E) approaches and identifying improvements. After reviewing current approaches, results were compiled and differences among engine companies were identified. Engine companies participating in the review included Allison/Rolls Royce, Honeywell, General Electric, Pratt & Whitney, Rolls Royce, plc., and Williams International. Current approaches are not necessarily applied uniformly among all engine manufacturers.

The work reported here is an overview of the general process and an example application from a recent, fairly typical HCF test program. Nonstandard methods of analysis and evaluation are utilized to highlight shortcomings in the current approach and to demonstrate potential improvements.

Process Overview

The process documented herein begins once hardware is available. This, therefore, assumes that the components that were manufactured have already met the required design criteria necessary for a part to move from the design phase to the verification phase. Early design studies are aimed at insuring that modal response frequencies do not coincide with known excitation sources. Computed natural frequencies are usually plotted on so-called Campbell diagrams (Fig. 1). The abscissa of this diagram is rotor speed. The ordinate is frequency in cycles per second or thousands of cycles per second (Hz, or KHz). Known excitation sources whose frequencies are functions of rotor speed are also plotted on this diagram. These types of excitations are typically generated by up or downstream pressure perturbations generated by adjacent blade rows. Intersection of a component natural frequency with a known excitation frequency generally produces a resonant response. Most design criteria seek to insure that these intersections are avoided at steady state rotor speeds.

Impedance Testing. A first step in the experimental process is a component test referred to as impedance testing. This approach is very cost effective since it does not require elaborate

instrumentation or excitation equipment. The test is used to determine modal frequencies for many or even all airfoils in a stage. Frequency verification is important since many HCF acceptance approaches are based upon insuring that potentially damaging modes are out of the engine running range. The tests also provide a quick assessment of the degree of frequency scatter as well as a first look at how well modal characteristics were captured analytically.

Holography Testing. Assuming experimental modal frequencies compare reasonably well with the analytical results, the next component test that is widely performed is holography testing. This optical test method produces images of mode shapes. Results are typically compared to contour plots of displacement from a finite element analysis. Comparison of analytical and experimental mode shapes is usually qualitative, that is, the images are checked to see if they “look” the same. Quantitative methods like the modal assurance criteria (MAC) are not typically used to assess mode shape agreement since holographic tests don’t generally produce displacement data. Only images are created. Methods like scanning laser vibrometry, which produce field measurements of surface velocity, could be used with the MAC to quantitatively validate analytical models.

Determination of Maximum Stress Locations. Standard approaches for identifying maximum modal stress locations can vary widely. Some manufacturers utilize a finite element model of a nominal part geometry exclusively to determine maximum relative modal stresses for each mode. Others use an entirely experimental process using stress pattern analysis by thermal emissions, or SPATE ([3]), and strain gages during a component test. Sometimes combinations of these approaches are used. Often the approach used depends on the hardware complexity. In all cases the

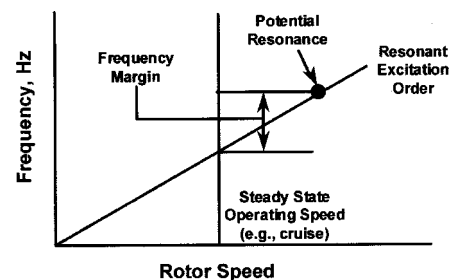


Fig. 1 Notional Campbell diagram

Contributed by the Materials, Manufacturing, and Metallurgy Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the MM&M Division, Oct. 2000; final revision received by the ASME Headquarters Feb. 2001. Editor: H. D. Nelson.

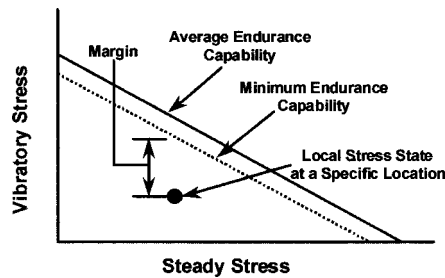


Fig. 2 Notional Goodman diagram

objective is the same; to identify locations of maximum modal, or vibratory, stress. Principal stresses, and sometimes VonMises stresses, are evaluated.

The information resulting from these experimental and/or analytical investigations is used to select preliminary strain gage locations that will be used in subsequent component and engine tests. Engine strain gages are generally not placed at the location of maximum modal stress so as to improve strain gage life. Also, since the allocation of strain gages to a particular part is often insufficient to permit placement of a strain gage at the maximum location for all the modes of interest, analysts must select locations that provide a reasonable measure for several modes. Methods used to arrive at these locations are most often based upon trial and error. All of these factors make it necessary to derive ratios from measurement locations to maximum stress locations for each mode of interest. These ratios are typically called "stress ratios."

The ratios are used to relate measured strains to maximum stresses. Consequently, any errors produced in this process can produce significant errors in subsequent tests. Sometimes, these ratios are adjusted for changes in mode shapes that may result from engine operation (e.g., centrifugal loading). Adjustments are performed using the finite element method despite the fact that the finite element method is typically only qualitatively validated.

Stress limits for each mode of concern are derived using the modal stress data, steady stress data from the finite element model at "critical" flight conditions, and fatigue data generated on material specimens or possibly on actual parts. Critical flight conditions are usually defined by company policy or engine specification. Data are plotted on a constant-life diagram called the Goodman diagram, Fig. 2.

This phase of engine test preparation can be the most time and cost intensive. Given the fact that this work must be completed prior to engine testing, significant schedule and cost pressures are often encountered increasing the potential for mistakes. This is an unfortunate consequence since the value of the subsequent engine tests depends heavily on this work.

Engine Testing. Once strain gage locations are selected and gages are applied, testing in component rigs or engines can proceed. Sea level engine tests are generally performed first to quantify modal responses. Sea level tests are less costly to perform than altitude tests which require a facility or an aircraft. The tradeoff for saving money is that data are only acquired at a single flight condition (sea level, static). The requirement for altitude testing depends largely on the engine mission requirements. In order to understand the influence of various operational variables on modal responses, testing under simulated altitude conditions at a ground test facility like the Arnold Engineering Development Center (AEDC), or in an aircraft during flight tests, is required.

Historically, the primary objective for structural evaluation of an engine has been to measure the maximum response a component will likely encounter in operational use—a deterministic approach. The search for the maximum response potentially spans many variables. Some modern, low aspect ratio airfoils can have as many as 60 modes below 30 KHz. Blade-to-blade variabilities

can be significant, and responses can vary with operational environment (flight condition). Deterministic approaches for identifying maximum responses are challenged, at best. For example, consider that for a single fan stage there may be, say, ten airfoils instrumented with, say, 20 strain gages to measure responses for up to 60 modes under 30 or more operational conditions. This test program would generate as many as 360,000 responses. Each of these responses would be ratio'd to the pretest identified critical location, and compared against a pretest-defined limit. From this, as few as one or two may actually be reported. Most are not even examined in any detail.

Methods used for processing and analyzing this large volume of data range widely among manufactures and test facilities. Some simply cull the data by online observation. If no obvious responses are noted during the test as displayed on an analog oscilloscope, no further processing or analysis is performed. At the other end of the spectrum, new tools are becoming available which permit all of the responses to be assessed and databased. Some of these tools will be referred to in subsequent sections.

Example Test Program

Test Description. A test program was recently completed at the AEDC which utilized all of the steps just described and generated a significant database of information. Three integrally bladed rotor (IBR) stages were instrumented with dynamic strain gages and tested at a variety of operational conditions. The rotor stages will be referred to as Rotor A, B, and C. Six different airfoils from Rotor A were instrumented at a single location. Twelve different airfoils on Rotor B were instrumented with up to two strain gages per airfoil. Locations for strain gage pairs were selected from four predefined locations. Total strain gage count for Rotor B was 24. Nine airfoils on Rotor C were instrumented with as many as two strain gages in a similar fashion as Rotor B. Locations for gage pairs used on Rotor C were selected from four pretest defined locations. Total strain gage count for Rotor C was 17. In all, 47 dynamic strain gages were used to generate vibratory response data. Strain gage locations and allocations are provided in Figs. 3–5.

The gages used during engine testing were also used during shaker table testing along with smaller, temporary strain gages affixed at critical locations identified by SPATE to derive the required relative stress ratios for most of the modes for each stage up to 20 KHz.

Data Processing and Analysis. Two-minute engine acceleration and/or deceleration ramps were run at 14 flight conditions with a variety of engine control schedules. These test variables, in addition to rotor speed, are subsequently referred to as *influence parameters*. Six modes from rotor A, 17 modes from Rotor B, and 14 modes from Rotor C responded at various rotor speeds during the test program. Only resonant responses were evaluated.

Dynamic data were processed using the AEDC developed computer assisted dynamic data monitoring and analysis system, or CADDMAS ([4]), and summarized using the newly developed DatWizard stress summary program ([5]). The DatWizard reviews the bulk data and databases peak responses for modes of interest. Modal stress ratios to the predefined critical locations were ap-

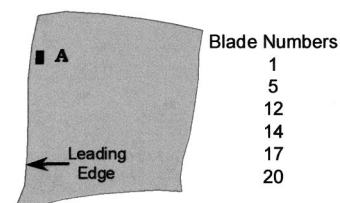


Fig. 3 Rotor A gage locations

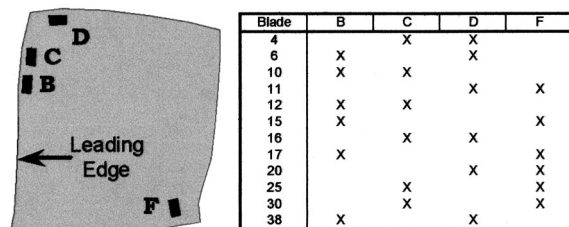


Fig. 4 Rotor B gage locations

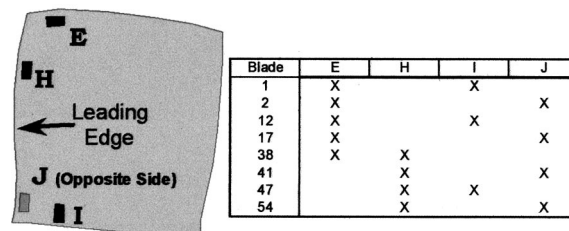


Fig. 5 Rotor C gage locations

plied to define the stress at critical locations and results were compared against pretest defined HCF limits. These tools were used to process and evaluate more than 9000 responses generated during the test program.

Statistical Analysis. Recognizing the probabilistic nature of HCF problems and, given that the engine community at large was developing a probabilistic HCF design system, the more than 9000 responses were evaluated statistically. Data were segregated by rotor stage, blade, mode, and engine order. After sorting the segregated data, median ranks were computed using the relation

$$\text{median rank} = i/(n+1),$$

where i is the data point for which a median rank is desired and n is the total number of data points for a given blade, mode, and engine order. This approach permitted the data to be statistically examined without assuming any particular statistical model. Results were presented graphically as plots of median rank versus normalized response amplitude at the predefined critical location. Response amplitudes were normalized to the response limit. Consequently, a value of 1.0 indicates that a given response reached the predetermined HCF limit. The resulting plots contained interesting characteristics that highlight several important deficiencies in the current HCF evaluation approach.

Analysis Results

After examining the database, statistical analysis was performed on 24 of the modes from the three rotor stages. Modes were selected that responded significantly during the test. Fundamental modes as well as highly complex modes were investigated. Four distinct characteristics are presented that address mode identification, instrumentation practices, and blade-to-blade variations. The data as presented also provides a means of defining what a "robust" blade design is.

Robust Characteristic. The first characteristic defined is the *robust characteristic*. By definition, a robust design is one that is tolerant to variations. Consequently, one would expect vibratory response amplitudes for a given mode and rotor stage to be invariant over a range of operational influences. Further, blade-to-blade variabilities should be small. Such a characteristic is presented in Fig. 6. Note that all the blades had similar responses over the range of influence parameters tested. Maximum responses were

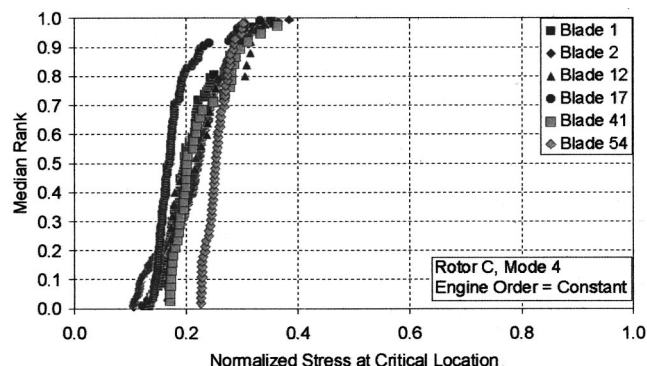


Fig. 6 Robust characteristic

below the limit value of 1.0. Individual blade characteristics are generally smooth; that is, there are no large discontinuities among the data points within a blade data set.

Nonrobust Characteristic. In contrast to the robust characteristic, an example of a *nonrobust characteristic* is presented in Fig. 7. Current data analysis methods would have identified a problem with this mode since the maximum response was above the limit value of 1.0. However, current approaches would probably only report this one data point on a Goodman diagram. Clearly, a better picture of the response characteristic can be derived from looking at all the response data.

Aside from the obvious fact that the maximum response amplitude is above the limit value (>1.0), several other interesting features are apparent. First, blade-to-blade variability is significant. Statistically, this may indicate that other blades are responding even higher than the instrumented ones. The design could be classified as sensitive to geometric variations and therefore not robust.

Notice also that the slope for some of the blades is low (larger range of stress). This indicates a sensitivity to influence parameters since each collection of data points were acquired over a large range of test conditions. It is statistically likely that testing under other conditions would produce higher responses.

Using today's criteria for determining whether response amplitudes are acceptable, this mode would be flagged as having exceeded predefined limits and would therefore be unacceptable. Some kind of corrective action would therefore be in order. However, this conclusion would be based almost exclusively on the fact that blade 17 measure responses in excess of predefined limits and would ignore the degree of blade-to-blade variability or the high sensitivity to the influence parameters. Indeed, had blade 17 not been instrumented, the modal response would be interpreted as acceptable.

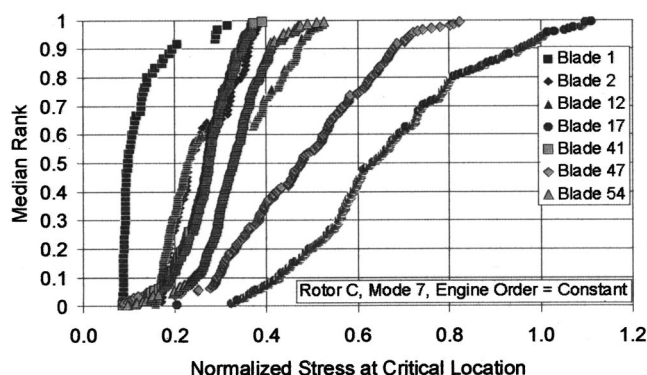


Fig. 7 Nonrobust characteristic

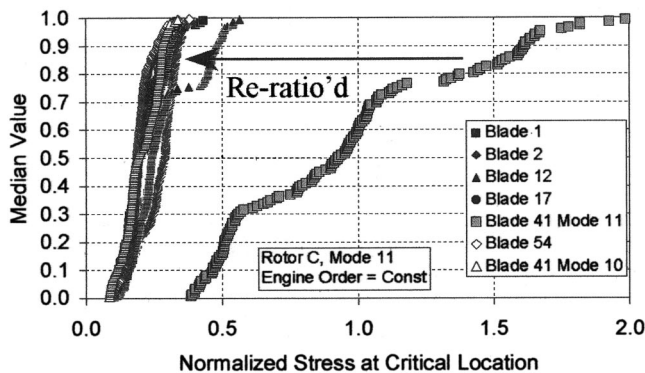


Fig. 8 Rogue blade characteristic

Rogue Blade Characteristic. Another example of what could be called a nonrobust response characteristic is presented in Fig. 8. This example clearly shows a single blade responding significantly different from all the others. This is the *rogue blade characteristic*. This type of behavior would probably be attributed to mistuning ([6,7]). Note that this response is significantly over the limit value. Consequently, it would be flagged as problematic and corrective action would be in order. The rogue blade also seems to be more sensitive to variations in influence parameters as indicated by the lower slope and increased range of response amplitude.

Interestingly, the mode shown in Fig. 8, mode 11, is very close in frequency to mode 10, calling into question whether the mode has been properly identified. If it is assumed that the mode has been incorrectly identified and the ratios to the critical location for mode 10 are applied instead of those for mode 11, the response characteristic is lower in amplitude (note the triangles—blade 41 mode 10 in the legend). Obviously, results would then be presented with other mode 10 data. Nonetheless, this observation highlights the criticality of identifying modes correctly. The relative strain ratios between the two gages applied to this blade were nearly identical for modes 10 and 11. Consequently the relative ratio between gages could not be used to help in mode identification. Ratios to the assumed critical location were markedly different, however. Today's approach typically utilizes only the response frequency to identify modes. Strain ratios are rarely used. Indeed, the fact that it is common practice to apply a single gage to a blade proves this point. This finding suggests that improvements in instrumentation protocols are needed. Also from this observation, note that changes in applied strain ratios can manifest themselves as slope changes in the response characteristic. Presentation of the data in this manner then, permits quick identification of potential instrumentation problems.

Disjointed Characteristic. Another characteristic present in the data set is referred to as the *disjointed characteristic*. It is evident in Fig. 8 most notably on blade 12. Blade 12 had two gages applied during the test identified as gage E and gage I. Independent examination of the data from each of these gages, Fig. 9, reveals that the higher response amplitudes on the composite response characteristic were derived from ratioing measured values from what is identified as gage I and the lower response amplitudes are from gage E. Separation of the two response populations from each of the gages shows that the variations to the influence parameters are the same as they should be (same slopes, Fig. 9). It also indicates that there is an inconsistency in pretest ratios to the critical location since obviously the strain at the critical location is a unique value.

In an attempt to understand the inconsistencies in the ratios to the critical location, strain ratios between the two gages were examined using calculated results from a zero RPM analysis and an analysis of the modes at approximately 85 percent of maximum

physical rotor speed. Mode 11 responds at approximately 85 percent maximum rotor speed. Gage strains were extracted from the finite element model output using the algorithm developed by Nichol [8] to insure comparisons between consistent units. Zero-speed ratios normalized to gage E agreed reasonably well with pretest experimental lab ratios (Fig. 10). However, note that the strain ratios for this mode are very different at 85 percent physical speed. Assuming these changes to be correct and applying the result to the measured data for gage I essentially duplicates gage E data for blade 12 in Fig. 9 (see gray diamonds and "Gage I reratio'd" in legend). It can therefore be reasonably concluded that the disjointed characteristic is a manifestation of strain ratio errors among multiple gages on a single part.

As a way of improving agreement between experimental and analytical strain ratios for gages E and I shown in Fig. 10, an assessment of gage placement uncertainty was conducted. The location and orientation of gages E and I were varied within what was assumed to be representative strain gage application tolerances. These tolerances were, $\Delta x, \Delta y = \pm 0.020$ inches in the plane of the gage in both the longitudinal and lateral directions, respectively, and $\Delta \theta = \pm 20$ deg rotation about the normal to the plane of the gage. Results are presented in Fig. 11. Note from the figure that "perfect" agreement between the pretest experimental results and the numerical results could be achieved if one allows for "small" variations in the actual location or orientation of the gages. Obviously, these "permitted" variations must be bounded by consideration of strain ratios from all modes. A method for doing this was documented by Sensmeier and Nichol [9].

Rogue Data Point Characteristic. A fourth and final response characteristic is presented in Fig. 12 which is perhaps the most difficult to handle of them all. It is the *rogue data point characteristic*. In-depth investigation of this response (calibration, data processing, test condition, etc.) revealed no discernable rea-

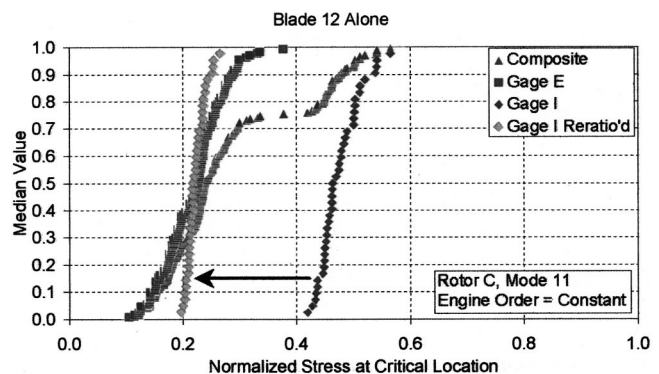


Fig. 9 Explanation of disjointed characteristic

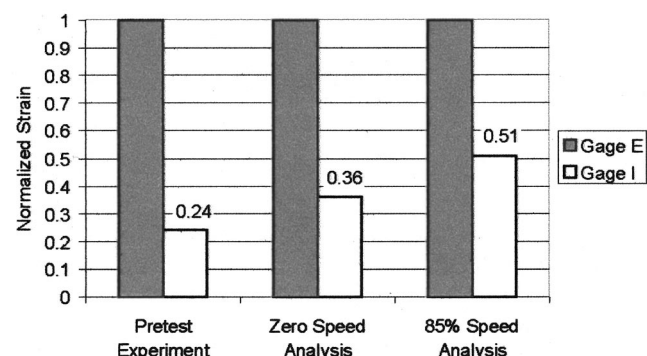


Fig. 10 Gage E and I strain ratios for mode 11

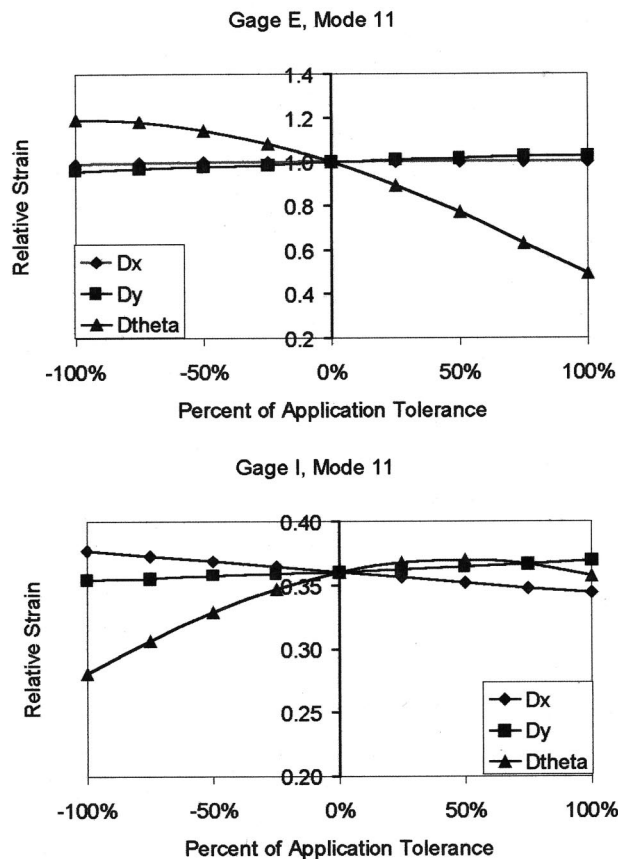


Fig. 11 Gage E and I application sensitivities

son that this data point would be any higher than any of the others. This characteristic is an excellent example of why unexpected HCF problems may continue to occur. Improvements in the T&E process can provide meaningful reductions, however.

Conclusions From the Assessment

An overview of some of the general findings from the assessment follow.

1 For vibratory response characterization, correct mode identification is critical. Modes must be correctly identified since to do otherwise will insure that incorrect ratios to critical stress locations are applied and that likewise incorrect values for the stress at the critical location will be inferred (given the current approach for determining stress at critical locations). Consequently, parts

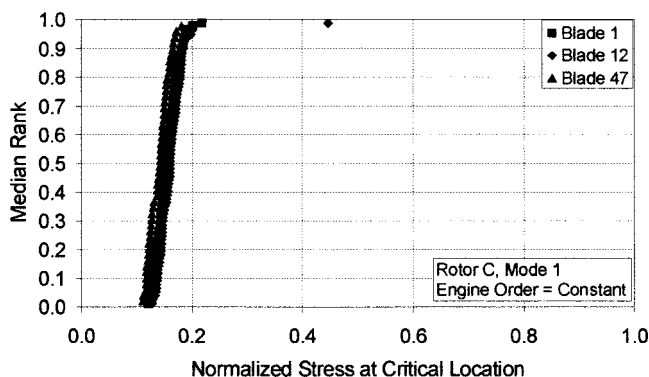


Fig. 12 Rogue data point characteristic

must be instrumented to insure that modes can indeed be identified by something other than frequency. This finding speaks to the need for an improved instrumentation protocol, reduced measurement uncertainty, and more efficient data processing.

2 Differences in mode shapes produced by variations in influence parameters (to include rotor speed) must be understood. These changes effectively produce changes in the ratios from measured locations to critical locations and produce the same error result as for(1). Note that this error can occur at the beginning of the test process in the determination of the base eigenvectors (mode shapes). If the modes change between the bench calibration test and the engine test because of boundary conditions or influence parameters, then the ratios are already in error.

3 Efforts must be made to identify real rogue blades and to quantify blade-to-blade, and engine-to-engine variations. This can be done statistically as shown in the preceding figures or via optical methods like NSMS [10, 11] provided the modes can be properly identified using these methods. Failure to identify modes correctly using optical methods results in the same errors as in(1) and (2) above. This is significant given that current optical approaches tend to rely on measurement of a single location on the blade, which does not permit displacement ratios to be obtained for correct mode identification. Also, most optical techniques have difficulty resolving synchronous response frequencies making correct mode identification even more difficult.

4 The test article must be exposed to either a statistically significant portion of all of the influence parameters or individual sensitivities to the influence parameters must be identified directly. Upon doing this, the maximum expected range of response amplitude on a per blade basis could be established.

5 HCF margin assessment must be conducted at all locations on the component. One of the key difficulties of the preceding analysis was the ratioing of measured results to an assumed minimum margin location. A better approach would be to provide a means of extrapolating or inferring the margins at all locations from the instrumented locations using a validated analytical model. The method would be able to handle changes in response vectors resulting from the influence parameters. A simple example of this is the ability to account for changes in mode shapes and the resulting strain ratios to the maximum stress location resulting from centrifugal effects. Additionally, this approach would provide assessment of all combinations of steady and dynamic stresses.

6 Efforts must be made to quantify the effects of measurement uncertainties. From this review, the most significant uncertainties relate to gage misplacement within application tolerances and inconsistent strain ratios to critical locations.

Recommendations

An overview of current turbine engine high-cycle fatigue test and evaluation methods has been presented. A recent test utilizing these methods has been assessed using supplemental approaches that have shown that there is potential for significant uncertainty or error in the determination of component HCF capability. This work establishes a baseline from which improvements in the HCF test and evaluation process can begin. Simply presenting all the data that is acquired as was done in the preceding sections is perhaps a good start. Processes that reduce or eliminate the reliance upon assumptions should be pursued. As an example, utilization of analytical models in conjunction with experimental results such as proposed by Nichol et al. [12], may be an alternative method for identifying the minimum HCF margin location on a given component. This approach requires that careful validation of analytical models be performed. The recommended method for this is scanning laser vibrometry and the modal assurance criteria. Once the model is sufficiently validated or otherwise modified, it can be used to select optimum strain gage locations, define regions with low fatigue margin, and to quantify the impact of influence parameters.

Acknowledgments

The author wishes to acknowledge the participation of the representatives from all the organizations represented on the HCF Test Protocol Working Group.

References

- [1] Heiser, W. H., and May, R., 1996, "U.S. Air Force Engine Management...A Cradle-to Grave Process," *Global Gas Turb. News*, **36**, No. 2, p. 9.
- [2] Thompson, D. E., and Griffin, J. T., 1999, "The National Turbine Engine High Cycle Fatigue Program," *Global Gas Turb. News*, **39**, No. 1, pp. 14–17.
- [3] Purcell, T. E., 1996, "Dynamic Stress Analysis of Gas Turbine Rotor Airfoils Using Thermoelastic Techniques," *Exp. Tech.*, May/June, pp. 9–13.
- [4] Tibbals, T. F., Bapty, T. A., and Abbot, B. A., 1994, "CADDMAS: A Real-Time Parallel System for Dynamic Data Analysis," ASME Paper No. 94-GT-194.
- [5] Tibbals, T. F., Sensmeier, M. D., Nichol, K. L., Stoner, A., and Biegl, C., 1999, "Model Integrated Testing for Durability Assessment of Turbine Engines," ITEA Conference Proceedings, University of Tennessee Space Institute, Tullahoma, TN, Oct.
- [6] Griffin, J. H., and Hoosac, T. H., 1984, "Model Development and Statistical Investigation of Turbine Blade Mistuning," *ASME J. Vib. Acoust., Stress Reliab. Des.*, **106**, pp. 204–210.
- [7] Ottarsson, G. S., Castanier, M. P., and Pierre, C., 1999, "A Reduced-Order Modeling Technique for Mistuned Bladed Disk," *Proceedings of the 35th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*.
- [8] Nichol, K. L., 1998, "Numerical Strain Gage Representation," *Proceedings of the 39th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, Long Beach, CA, Apr.
- [9] Sensmeier, M. D., and Nichol, K. L., 1998, "Minimizing Vibratory Strain Measurement Error," *Proceedings of the 43rd ASME Gas Turbine and Aeroengine Congress*, Stockholm, Sweden, June.
- [10] Jones, H. T., 1985, "Development of Noninterference Technique for Measuring Turbine Engine Rotor Blade Stresses," AIAA Paper 85-1472.
- [11] Watkins, A. B., Robinson, W. W., and Chi, R. M., 1985, "Noncontact Vibration Measurement and Analysis," AIAA Paper No. 85-1473.
- [12] Nichol, K. L., Sensmeier, M. D., and Tibbals, T. F., 1999, "Assessment of Turbine Engine Structural Integrity Using the Structural Dynamic Response Analysis Code (SDRAC)," ASME Paper No. 99-GT-410.

S.-M. Wang

Division of Theoretical Mechanics,
School of Science,
Beijing University of Aeronautics and
Astronautics,
Beijing 100083, People's Republic of China

Q.-S. Lu

Division of Theoretical Mechanics,
School of Science,
Beijing University of Aeronautics and
Astronautics,
Beijing 100083, People's Republic of China
and
Department of Mathematical Sciences,
Brunel University,
Uxbridge, Middlesex UB8 3PH, UK

E. H. Twizell¹

Department of Mathematical Sciences,
Brunel University,
Uxbridge Middlesex UB8 3PH, UK
e-mail: E.H.Twizell@brunel.ac.uk

Reducing Lateral Vibration of a Rotor Passing Through Critical Speeds by Phase Modulating

A method is presented to reduce the lateral vibration amplitude of an unbalanced rotor accelerating or decelerating through its critical speed, by means of modulating the response phase with varying acceleration rate. Theoretical and numerical analysis on the amplitude and the phase characteristics of lateral vibration of a rotor model are made. Numerical results show that when the rotor passes through its critical speed with a modulated phase, the response amplitude can be reduced by 20% or so, compared with the nonmodulated (constant) acceleration case. [DOI: 10.1115/1.1581893]

1 Introduction

The large amplitude at critical speeds is one of the most serious concerns of high-speed rotating machines. It may cause tip-rubs of rotor and guard in rotating machines as well as increased bearing reactions, [1–3], and the situation can be catastrophic. In order to avoid the damage due to rubbing, nominal tip clearances are specified well in excess of those necessities for steady-state supercritical operation. This degrades the turbomachine's aerodynamic performance. In particular, large clearances create low efficiencies for both turbines and compressors. Therefore, understanding the mechanism of the lateral vibration of a rotor passing through critical speeds and finding methods to reduce the critical amplitude are important.

In industry practices, the general method to avoid large amplitude resonance at critical speeds is to use a larger acceleration when passing through the critical speed, [4]. This measure is limited by the driving power. Millsaps and Reed [5] suggested a method for reducing the lateral vibration response of an unbalanced rotor accelerating (or decelerating) through its first critical speed by using a variable acceleration rate. According to the amplitude-speed characteristic with various acceleration rates, they designed an acceleration schedule. An initial large acceleration is used to pass the critical speed, followed by a smaller acceleration, expecting to reduce the maximum resonance amplitude. This is a good idea, but they only considered the action of the variation of acceleration itself, neglecting the effect of the corresponding variation of phase. As a result, the transient motion after acceleration switch may be different from the predicted result in Fig. 7 of [5]. Indeed, the amplitude behavior is strongly dependent on the phase at the position of acceleration switch. Therefore, it is important to analyze the phase relation between the inertial resultant force and the rotor response in the process of passing through the critical speed and to schedule the acceleration by means of phase modulation for reducing lateral vibration.

In this paper, taking into account the amplitude and the phase

characteristics of lateral vibration of a rotor passing through critical speeds, a more efficient method of scheduling acceleration by means of phase modulating is presented to reduce the maximum amplitude of vibration during accelerating or decelerating.

2 Mechanical Model

Figure 1 shows the model for an unbalanced rotor consists of a simple supported bent shaft and a mass eccentric rigid disk, and the following assumptions are made:

- The shape of the shaft is always a planar curve and the mass of the shaft is neglected.
- The axial and torsional vibrations are negligible.
- The operating speed is below the second critical speed of the rotor.
- The gyroscopic effect of the disk can be considered on account of the bending of the shaft.
- The strain energy of the shaft is given by

$$V = \frac{1}{2} k \rho_m^2 \quad (1)$$

where k is the equivalent stiffness of the shaft and ρ_m is the shaft deflection at the middle span.

The disk motion consists of two parts: the translation motion and the gyroscopic motion. The kinetic energy T of the disk is expressed as

$$\begin{aligned} T = & \frac{1}{2} m [\rho_d'^2 + \rho_d^2 \psi'^2 + e^2 \varphi'^2 + 2e \rho_d' \varphi' \sin(\psi - \varphi) \\ & + 2e \rho_d \varphi' \psi' \cos(\psi - \varphi)] + \frac{1}{2} \bar{I}_p [\psi' \cos \theta + (\varphi' - \psi')]^2 \\ & + \frac{1}{2} \bar{I}_d [\psi'^2 \sin^2 \theta + \theta'^2], \end{aligned} \quad (2)$$

where the prime denotes the derivative with respect to time t ; ρ_d is the shaft deflection at the mounting position of the disk; ψ is the angle between the shaft deflection plane and xz -plane; e is the distance from the mass center of the disk to the shaft center; φ is absolute rotation angle of line OO' ; I_p , I_d are diametric and polar moments of inertia, respectively; θ is the angle between the disk normal line and z -axis (i.e. the disc nutation angle) shown in Fig. 1.

For simplicity, suppose that the shaft deflection is expressed as

$$\rho(z, t) = \hat{q}(t) \sin(\pi z / 2l),$$

¹To whom correspondence should be addressed.

Contributed by the Structures and Dynamics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the S&D Division Oct. 1, 2001; final revision received by the ASME Headquarters Oct. 28, 2002. Associate Editor: M. Mignolet.

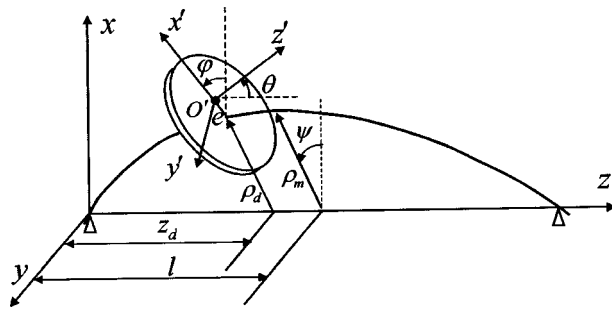


Fig. 1 Mechanical model of an unbalanced rotor

where $\sin(\pi z/2l)$ is the first model of the shaft; $\hat{q}(t)$ is the time variable of the shaft deflection; the length of the shaft is $2l$. Then we have

$$\rho_m = \hat{q}(t), \quad \rho_d = \hat{q}(t) \sin(\pi z_d/2l),$$

$$\tan \theta = \hat{q}'(t) (\pi/2l) \cos(\pi z_d/2l),$$

where z_d is the z -coordinate of the disk position. For small bending deflection, the following expression can be obtained:

$$\theta \approx \hat{q}'(t) (\pi/2l) \cos(\pi z_d/2l), \quad \theta' \approx \hat{q}''(t) (\pi/2l) \cos(\pi z_d/2l).$$

Introduce the dimensionless variables and parameters:

$$q = \rho_d/e = (\hat{q}(t)/e) \sin(\pi z_d/2l), \quad \lambda = (\pi e/2l) \cot(\pi z_d/2l),$$

$$I_p = \bar{I}_p/me^2, \quad I_d = \bar{I}_d/me^2,$$

$$\omega_0^2 = k/[m(\sin \pi z_d/2l)^2], \quad \delta = c/(m\omega_0), \quad \tau = \omega_0 t.$$

By means of the Ostrogradsky-Hamilton principle, the dimensionless equations of the rotor motion can be written as

$$(1 + \lambda^2 I_d) \ddot{q} - \dot{\psi}^2 q + \delta \dot{q} + q + \ddot{\varphi} \sin(\psi - \varphi)$$

$$- \dot{\varphi}^2 \cos(\psi - \varphi) + \lambda I_p (\dot{\psi} (\cos \lambda q - 1) + \dot{\varphi}) \dot{\psi} \sin \lambda q$$

$$+ \lambda I_d \dot{\psi}^2 \sin \lambda q \cos \lambda q = 0 \quad (3)$$

$$q^2 \ddot{\psi} + 2q\dot{q}\dot{\psi} + \delta q^2 \dot{\psi} + q\ddot{\varphi} \cos(\psi - \varphi) + q\dot{\varphi}^2 \sin(\psi - \varphi)$$

$$+ 2\lambda I_d \dot{q}\dot{\psi} \sin \lambda q \cos \lambda q + I_d \ddot{\psi} \sin^2 \lambda q$$

$$+ I_p [\ddot{\psi} (\cos \lambda q - 1) + \ddot{\varphi} - \lambda \dot{\psi} \dot{q} \sin \lambda q] (\cos \lambda q - 1)$$

$$- \lambda I_p \dot{q} [\dot{\psi} (\cos \lambda q - 1) + \dot{\varphi}] \sin \lambda q = 0 \quad (4)$$

where the dot denotes the derivative with respect to the dimensionless time τ .

Introducing the following transformations

$$F_I = \sqrt{\dot{\varphi}^2 + \dot{\psi}^4}, \quad \beta = \arctan(\ddot{\varphi}/\dot{\varphi}^2),$$

we have

$$\ddot{\varphi} \sin(\psi - \varphi) - \dot{\varphi}^2 \cos(\psi - \varphi) = -F_I \cos(\varphi - \psi - \beta),$$

$$\ddot{\varphi} \cos(\psi - \varphi) + \dot{\varphi}^2 \sin(\psi - \varphi) = -F_I \sin(\varphi - \psi - \beta).$$

Substituting the above relations into Eqs. (3) and (4), we get

$$\ddot{q} - \dot{\psi}^2 q + \delta \dot{q} + q - F_I \cos(\varphi - \psi - \beta) + \lambda^2 I_d \ddot{q}$$

$$+ \lambda^2 q I_p \dot{\psi} (\dot{\varphi} - \frac{1}{2} \dot{\psi} \lambda^2 q^2) + \lambda^2 I_d q \dot{\psi}^2 = 0, \quad (5)$$

$$q\ddot{\psi} + 2\dot{q}\dot{\psi} + \delta q\dot{\psi} - F_I \sin(\varphi - \psi - \beta) + 2\lambda^2 I_d \dot{q}\dot{\psi} + I_d \lambda^2 q \ddot{\psi}$$

$$- \frac{1}{2} \lambda^2 q I_p (\ddot{\varphi} - \frac{1}{2} \lambda^2 q^2 \ddot{\psi} - \lambda^2 q \dot{\psi} \dot{q}) - \lambda^2 I_p \dot{q} (\dot{\varphi} - \frac{1}{2} \dot{\psi} \lambda^2 q^2) = 0, \quad (6)$$

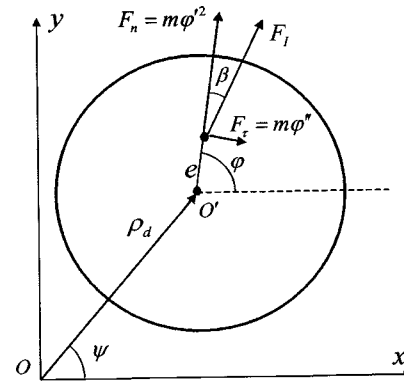


Fig. 2 Geometric angles and the inertia forces

where the angles φ , ψ , β are shown in Fig. 2. In Fig. 2, O is the geometric center of the stator, ρ_d represents the displacement of the disk geometric center O' , e is eccentricity of the disk, F_n and F_t are the normal and tangential inertia forces, respectively, and F_I is the resultant inertia force.

The gyroscopic motion of the disk is described by three Euler angles: θ (nutation), ψ (precession) and $\varphi - \psi$ (rotation), respectively. When the disk is mounted at the middle span of the shaft, the disk has no gyroscopic motion corresponding to $\theta = 0$.

3 Energy Analysis

To investigate the characteristic of the lateral variation amplitude of the rotor during acceleration, the dimensionless mechanical energy is introduced:

$$H = \frac{1}{2} (\dot{q}^2 + q^2 \dot{\psi}^2 + q^2). \quad (7)$$

This expression contains the dimensionless stress energy of the shaft, $1/2q^2$, and the translation motion kinetic energy of the disk, namely $1/2(\dot{q}^2 + q^2 \dot{\psi}^2)$. Both of them reflect the variation of the amplitude. Differentiating Eq. (7) with respect to the nondimensional time τ and taking account of Eqs. (5) and (6), we have

$$\dot{H} = \dot{q}\ddot{q} + q\dot{q}\dot{\psi}^2 + q^2\dot{\psi}\ddot{\psi} + q\dot{q}$$

$$= F_I [\dot{q} \cos(\varphi - \psi - \beta) + q\dot{\psi} \sin(\varphi - \psi - \beta)]$$

$$- \delta(\dot{q}^2 + q^2 \dot{\psi}^2) + \dot{H}, \quad (8)$$

where

$$\dot{H} = \lambda^2 [-\dot{q}(I_d \ddot{q} + q I_p \dot{\psi} (\dot{\varphi} - \frac{1}{2} \dot{\psi} \lambda^2 q^2) + I_d q \dot{\psi}^2)$$

$$- q \dot{\psi} (2 I_d \dot{q} \dot{\psi} + I_d q \ddot{\psi} - \frac{1}{2} q I_p (\ddot{\varphi} - \frac{1}{2} \lambda^2 q^2 \ddot{\psi} - \lambda^2 q \dot{\psi} \dot{q})$$

$$- I_p \dot{q} (\dot{\varphi} - \frac{1}{2} \dot{\psi} \lambda^2 q^2))]. \quad (9)$$

Taking the following transformations

$$E = \sqrt{\dot{q}^2 + q^2 \dot{\psi}^2}, \quad \alpha = \arctan(q\dot{\psi}/\dot{q}), \quad \gamma = \alpha - (\varphi - \psi - \beta),$$

where γ is the phase difference between the velocity phase and the phase of the resultant inertia force, using the above transformations, the expression (8) can be rewritten as

$$\dot{H} = E(F_I \cos \gamma - \delta E) + \dot{H}. \quad (10)$$

It is seen from Eq. (9) that $\dot{H} = O(\lambda^2)$. Therefore, for small λ (that is, small disk bending), the variation rate of the energy $\dot{H}$ is mainly determined by the first term of Eq. (10), in which the phase γ plays an important role. In this case, if $-\pi/2 < \gamma < \pi/2$, $\dot{H}$ may be either positive or negative; while if $\pi/2 < \gamma < 3\pi/2$, $\dot{H}$ must be negative.

4 Amplitude Characteristic in Accelerating Process

In order to investigate the relationship between the amplitude and the phase γ , Eqs. (5) and (6) were integrated numerically by means of MATHEMATICA 4.0 under the following parameter values: $\delta=0.04$, $\lambda=0$ and the angular acceleration $\ddot{\phi}=0.01$ and 0.001 , respectively. The numerical results are shown in Fig. 3. Fig. 3(a) gives the amplitude characteristics under different constant accelerations, where q is the vibration amplitude. Because of the effect of damping, the maximum amplitude occurs a little after the rotation speed passes through the lateral natural frequency of the shaft (i.e., $\dot{\phi}=1$); Fig. 3(b) shows the phase characteristics during the accelerating process, where the vertical axis is the sine of the phase γ . The variation rate of vibration energy $\dot{H}$ is given in Fig. 3(c). Referring to Figs. 3(a) and 3(c) it can be seen that the amplitude increases when $\dot{H}>0$, while the amplitude decreases when $\dot{H}<0$. This fact is evident because the vibration energy increasing or decreasing leads to the vibration amplitude increasing or decreasing.

Furthermore, it is seen from Fig. 3(a) that the maximum vibration amplitude depends on the magnitude of the acceleration when starting the rotor from the static state to the operating state. The larger the angular acceleration, the smaller the maximum amplitude is, and the later the maximum amplitude occurs. In this figure, two amplitude characteristic curves corresponding to $\ddot{\phi}=0.01$ and 0.001 , respectively, have an intersection point P . Millsaps and Reed [5] suggested that if at first the rotor accelerated with a larger acceleration to the point P and then changed the acceleration to a smaller one, a smaller vibration amplitude could be expected. According to this acceleration scheduling idea, a nu-

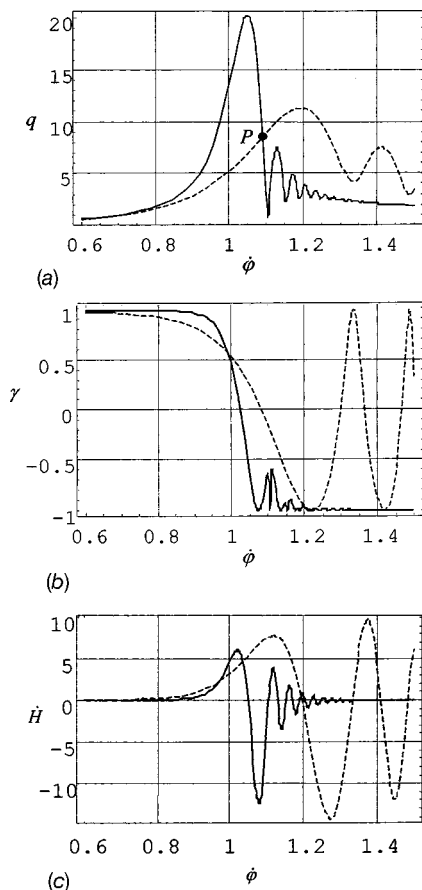


Fig. 3 Vibration characteristics under different constant accelerations, the solid curve is for $\ddot{\phi}=0.01$ and the dashed curve is for $\ddot{\phi}=0.001$. ($\lambda=0$, $\delta=0.04$). (a) Amplitude characteristic, (b) phase characteristic, (c) energy characteristic.

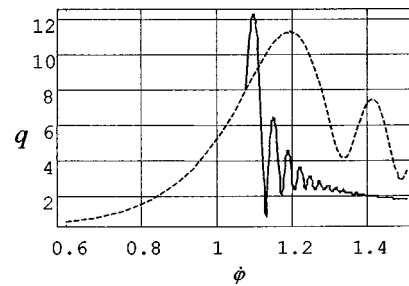


Fig. 4 Amplitude characteristic with acceleration switching, the dashed curve is for $\ddot{\phi}=0.01$ and the solid curve is for $\ddot{\phi}=0.001$ after acceleration switching

merical simulation was made. Accelerating the rotor with an angular acceleration $\ddot{\phi}=0.01$ to the point P (about $\dot{\phi}=1.08$) and then changing to $\ddot{\phi}=0.001$, the numerical result of the amplitude is shown in Fig. 4.

Indeed, the amplitude will continue to increase smoothly for a while just after the acceleration switching with the maximum amplitude larger than that in the case of constant angular acceleration $\ddot{\phi}=0.01$, and then it begins to decrease. Therefore, the simulation result is not like the predication shown in Figs. 5 and 6 of [5], in which the vibration amplitude drops suddenly just after the acceleration switching and the amplitude characteristic curve is nonsmooth at the point P . The reason for this is as follows. It can be seen from Figs. 3(b) and 3(c) that the point P corresponds to $\sin \gamma \approx 0.5$ and $\dot{H} \approx 7 > 0$, so the vibration amplitude should keep on increasing continuously near the point P as shown in Fig. 4. In order to reduce the maximum amplitude of lateral vibration of the rotor more efficiently, a phase modulating scheme is presented in this paper, instead of the acceleration scheduling scheme in [5], in which only the magnitude of the angular acceleration is considered.

5 Phase Modulating

From the above discussion, it is seen that the variation of vibration amplitude is determined by the variation rate of vibration energy $\dot{H}$. Taking account of the fact that the variation rate of vibration energy $\dot{H}$ is closely dependant on the phase γ from Eq. (10), it is possible to gradually adjust the phase to indirectly reduce the maximum vibration amplitude. A more effective way for controlling the phase γ is to change both the magnitude and the direction of angular acceleration.

An example is given to illustrate the idea of phase modulating scheme. At first, the rotor is accelerated up to a near-critical state (for example, $\dot{\phi}=0.948$) with an angular acceleration $\ddot{\phi}=0.01$, the variation of phase γ is shown by the curve a of Fig. 7(a). Then reverse the direction of angular acceleration by changing it to $\ddot{\phi}=-0.001$ in order to modulate the phase. Here the reverse acceleration is called 'modulating acceleration', and its magnitude should be adequately chosen. In our numerical simulation, the magnitude of the modulating acceleration is taken as one tenth of the original acceleration. After the angular velocity goes back to an adequate value (for example, $\dot{\phi}=0.894$), the phase γ approaches 2.5 as shown by the curve b , while it is 1.02 at the original state $\dot{\phi}=0.894$ before phase modulating as shown by curve a . The corresponding energy characteristic with phase modulating is shown in Fig. 7(b). It can be seen that $\dot{H} \approx 0.9$ at the original state $\dot{\phi}=0.894$ before phase modulating, but $\dot{H} \approx -3.2$ at the same state after phase modulating. This means that either continuing to decelerate or to accelerate the rotor again at this moment, the amplitude will keep on decreasing for a while. Therefore, it is obvious that the above schedule can modulate the phase from an amplitude-increasing phase to an amplitude-decreasing one, so that the maximum amplitude is reduced.

6 Examples

In what follows some numerical simulation results for Eqs. (5) and (6) are given and analyzed in consideration of phase modulating control. Both the case that the rotor is accelerated from the static state to the operating state (i.e., starting up) and the reverse case (i.e., shutting down) are investigated under phase modulating, where the effect of the gyroscopic motion of rotor can be taken into account if the parameter $\lambda \neq 0$.

6.1 Starting Up Case With $\lambda=0$. Suppose that the rotor is speeding up from the static state to the operating state $\dot{\phi}=1.5$. The disk is mounted at the middle span of the shaft, and then it has no gyroscopic motion with the parameter $\lambda=0$. A constant angular acceleration $\ddot{\phi}=0.01$ is taken to accelerate the rotor. The amplitude characteristic without phase modulating is shown in Fig. 5(c) by the dotted curve. Now consider the rotor motion with phase modulating. Changing the acceleration to $\ddot{\phi}=-0.001$ at the state $\dot{\phi}=0.948$ to decelerate the rotor to the state $\dot{\phi}=0.894$, the phase γ in this “to and fro” process is shown in Fig. 5(a) by the curves *a* and *b*, respectively. The variation rate of vibration energy $\dot{H}$ is shown in Fig. 5(b) by the curves *a* and *b*, respectively. After that, the rotor is accelerated with $\ddot{\phi}=0.01$ again. In this final stage, the phase γ is shown in Fig. 5(a) by the curve *c* and the variation rate of vibration energy $\dot{H}$ is shown in Fig. 5(b) by the curve *c*. The solid curve in Fig. 5(c) shows the amplitude characteristic with phase modulating. The maximum amplitude can be reduced by about 25%; this is seen by comparing the amplitude curves with and without phase modulating.

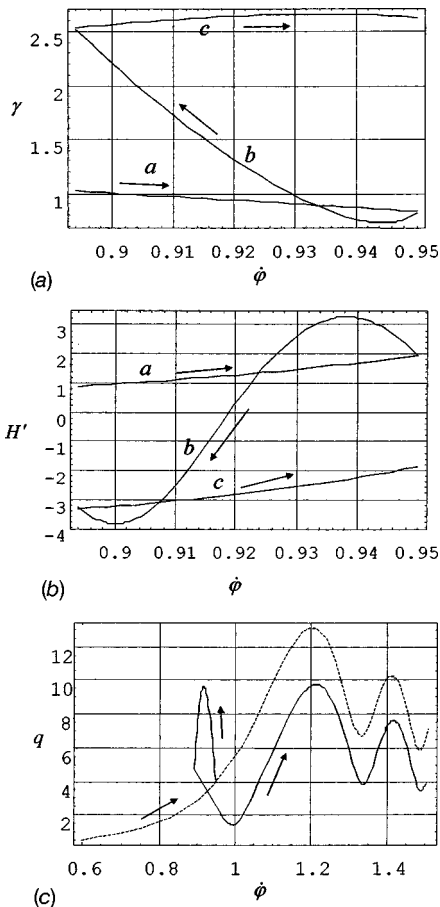


Fig. 5 Phase, energy, and amplitude characteristics with phase modulating ($\lambda=0$, $\delta=0.02$). (a) Phase characteristic; (b) energy characteristics; (c) amplitude characteristic.

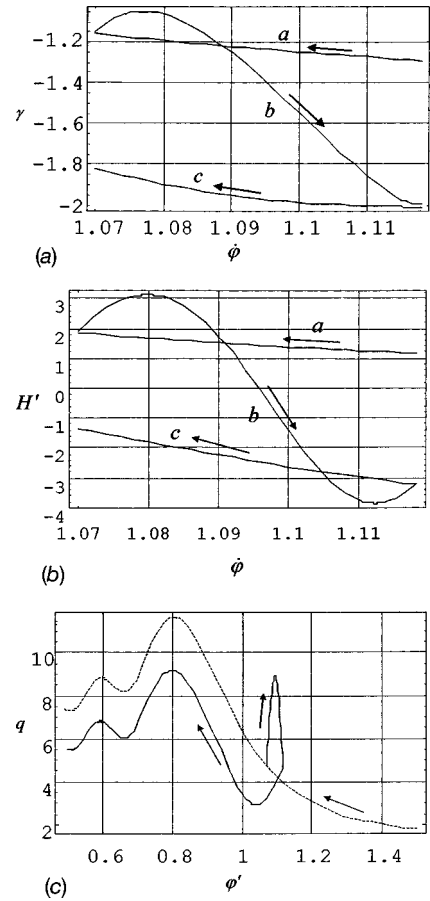


Fig. 6 Phase, energy, and amplitude characteristics with phase modulating ($\lambda=0$, $\delta=0.02$). (a) Phase characteristic; (b) energy characteristics; (c) amplitude characteristic.

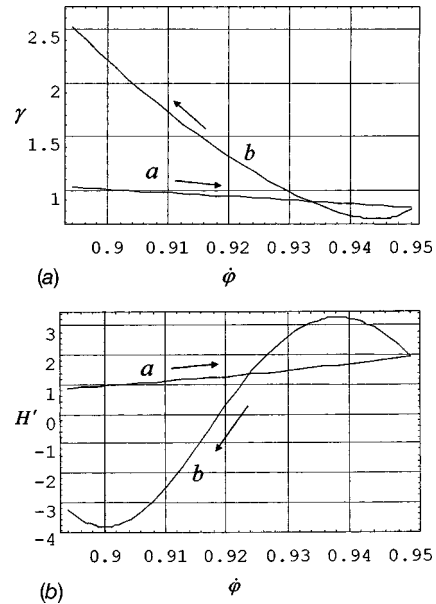


Fig. 7 Phase and energy characteristics ($\lambda=0$, $\delta=0.02$). (a) Phase characteristics: *a* $\dot{\phi}=0.01$; *b* $\dot{\phi}=0.001$. (b) Energy characteristics: *a* $\dot{\phi}=0.01$; *b* $\dot{\phi}=0.001$.

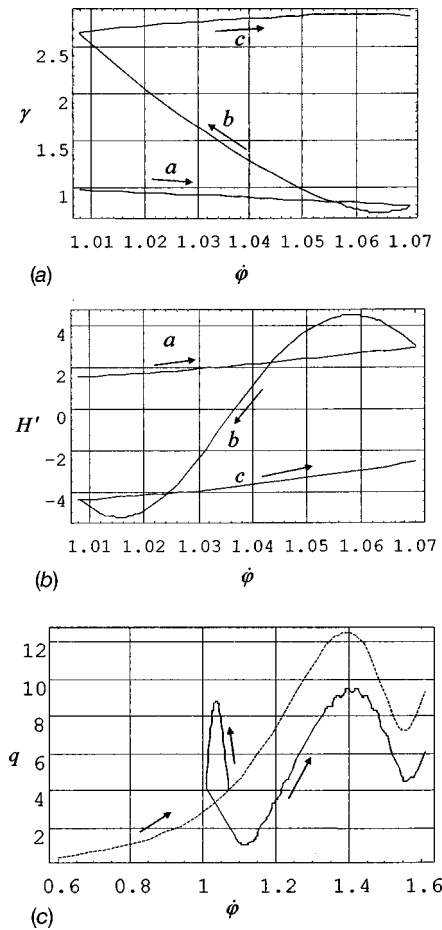


Fig. 8 Phase and energy as well as amplitude characteristics with phase modulating. ($\delta=0.02$, $\lambda=0.002$, $I_p=2I_d=250000$). (a) Phase characteristic; (b) energy characteristics; (c) amplitude characteristics.

6.2 Shutting Down Case With $\lambda=0$. Suppose that the rotor is slowing down from $\dot{\varphi}=1.5$ to the static state and the disk is also mounted at the middle span of the shaft. The amplitude characteristic with a constant angular acceleration $\ddot{\varphi}=-0.01$ is shown by the dotted curve in Fig. 6(c). Changing the acceleration to $\ddot{\varphi}=0.001$ at the state $\dot{\varphi}=1.07$, to accelerate the rotor to the state $\dot{\varphi}=1.118$, the phase γ in this “to and fro” process is shown in Fig. 6(a) by the curves *a* and *b*, respectively. The variation rate of vibration energy $\dot{H}$ is shown in Fig. 6(b) by the curves *a* and *b*, respectively. After that, the rotor is decelerated with $\ddot{\varphi}=-0.01$ again. The phase γ is shown in Fig. 6(a) by the curve *c* and the variation rate of vibration energy $\dot{H}$ is shown in Fig. 6(b) by the curve *c*. The solid curve in Fig. 6(c) shows the amplitude characteristic with phase modulating. The maximum amplitude is reduced about 20% by phase modulating.

6.3 Starting Up Case With Gyroscopic Motion ($\lambda \neq 0$). When the disk is mounted at a place other than the middle span of the shaft, there exists a gyroscopic motion of the disk along with the shaft vibration. Using the following relations

$$\theta \cong \hat{q}(t) (\pi/2l) \cos(\pi z_d/2l),$$

$$q = \rho_d/e = (\hat{q}(t)/e) \sin(\pi z_d/2l), \quad \lambda = (\pi e/2l) \cot(\pi z_d/2l),$$

the dimensionless nutation angle can be expressed as $\theta = \lambda q$. This means that the nutation amplitude of the disk depends on its mounting position and the vibration amplitude of the shaft.

A constant angular acceleration $\ddot{\varphi}=0.01$ is taken now to accelerate the rotor from the static state to the operating state $\dot{\varphi}=1.5$. The amplitude curve is shown in Fig. 8(c) by the dotted curve. Comparing with that in the case (1), the maximum amplitude occurs later because of the gyroscopic effect of the disk. Now consider the rotor motion with gyroscopic effect and phase modulating. If changing the acceleration to $\ddot{\varphi}=-0.001$ at the state $\dot{\varphi}=1.07$ to decelerate the rotor to the state $\dot{\varphi}=1.008$, the phase γ is shown in Fig. 8(a) by the curves *a* and *b*, respectively. The variation rate of vibration energy $\dot{H}$ is shown in Fig. 8(b) by the curves *a* and *b*, respectively. After that, the rotor is accelerated with $\ddot{\varphi}=0.01$ again. The phase γ is shown in Fig. 8(a) by the curve *c* and the variation rate of vibration energy $\dot{H}$ is shown in Fig. 8(b) by the curve *c*. The solid curve in Fig. 8(c) shows the amplitude characteristic with phase modulating. It can be seen that the maximum amplitude is reduced by about 20% by phase modulating, in which the gyroscopic motion of the rotor is taken into account.

7 Summary

This work is a theoretical and numerical analysis on the phase modulating method in rotor vibration control. Through a complete analysis on the amplitude-speed and the phase-speed characteristics as well as their relations, a phase modulating scheme by acceleration scheduling is presented to reduce the critical vibration amplitude of rotors. This method does not need complicated auxiliary equipment, and it can also be effective in reducing the maximum vibration amplitude in other vibration systems with critical speeds. Further experimental confirmation for this method is still needed in laboratories and engineering practices.

A possible way for phase modulating can be taken according to the variation of vibrating energy H . It can be seen from the above examples that the curves of $\dot{H}$ vary in an “S” shape during phase modulating. The phase modulating schedule can be arranged as follows: First, choose an appropriate angular acceleration to drive the rotor to the first threshold state with $\dot{H}=\dot{H}_0>0$ near the critical speed. Secondly, reverse the direction of angular acceleration and adjust its magnitude to modulate the phase until the second threshold state with $\dot{H}=\dot{H}_1<0$ and $\dot{H}_1>0$. Finally, use the original angular acceleration to drive the rotor passing through the critical state. In order to get $\dot{H}$ expressed as Eq. (8) in the process of vibration control, the variables involved in H can be measured.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China under Grant No. 19990510 and 10172011, the Science School of Beijing University of Aero-Astronautics, the Doctoral Education Foundation of China and The Royal Society of London.

Nomenclature

- c = damping (Ns/m)
- e = distance from the center of mass of the disk to the shaft center (i.e., eccentricity) (m)
- E = dimensionless kinetic energy T of the disk
- H = dimensionless mechanical energy of the disk translation motion
- $\bar{I}_p, \bar{I}_d$ = polar and diameter momentum of inertia of the disk, respectively (kg·m²)
- $I_p = \bar{I}_p/me^2, I_d = \bar{I}_d/me^2$ = dimensionless polar and diameter momentum of inertia of the disk, respectively
- k = equivalent stiffness of the shaft (N/m)

m = disk mass (kg)
 $\hat{q}(t)$ = time variable of shaft deflection (m)
 $q = \rho_d / e$ = dimensionless time part variable of shaft deflection at the mounting position of disk
 t = time (s)
 V = elastic potential energy of the shaft (Nm)
 $\delta = c / m \omega_0$ = dimensionless damping
 φ = absolute rotation angle of eccentric mass (rad)
 $\gamma = \alpha - (\varphi - \psi - \beta)$ = phase angle (rad)
 ω_0 = natural bending frequency of the shaft (rad/s)
 ψ = precession (or whirling) angle (rad)
 ρ_m = shaft deflection displacement at the middle span (m)
 ρ_d = shaft deflection displacement at the mounting position of disk (m)

θ = bending angle of the shaft at the mounting position of disk (nutation angle) (rad)
 $\tau = \omega_0 t$ = dimensionless time
 $Oxyz$ = absolute coordinates
 $O'x'y'z'$ = body coordinates of the disk

References

- [1] Choy, F. K., and Padovan, J., 1987, "Nonlinear Transient Analysis of Rotor-Casing Rub Events," *J. Sound Vib.*, **113**, pp. 529–545.
- [2] Li, G. X., and Paidoussis, M. P., 1994, "Impact Phenomena of Rotor-Casing Dynamical Systems," *Nonlinear Dyn.*, **5**, pp. 53–70.
- [3] Yanabe, S., Kanneko, S., and Shibata, N., 1998, "Rotor Vibration due to Collision With Annular Guard During Passage Through Critical Speed," *ASME J. Vibr. Acoust.*, **120**, pp. 549–552.
- [4] Fearn, R. L., and Millsaps, K., 1967, "Constant Acceleration of an Undamped Simple Vibrator Through Resonance," *J. of Royal Aero. Soc.*, Aug, pp. 567–573.
- [5] Millsaps, K. T., and Reed, G. L., 1998, "Reducing Lateral Vibrations of a Rotor Passing Through Critical Speeds by Acceleration Scheduling," *ASME J. Eng. Gas Turbines Power*, **120**, pp. 615–620.

P. Marugabandhu

J. H. Griffin

Department of Mechanical Engineering,
Carnegie Mellon University,
Pittsburgh, PA 15213

A Reduced-Order Model for Evaluating the Effect of Rotational Speed on the Natural Frequencies and Mode Shapes of Blades

A reduced-order model has been developed that can be used to accurately and quickly calculate the changes in the natural frequencies and mode shapes of a blade that are caused by centrifugal stiffening. It has been corroborated by comparisons with finite element analyses of a cantilevered tapered plate and with frequencies from a low aspect ratio fan blade. [DOI: 10.1115/1.1448327]

1 Introduction

When a structure has two modes with closely spaced natural frequencies, their corresponding mode shapes can interact if the structure is modified. By interact, we mean that the mode shapes of the new modes can be expressed as a linear combination of the original modes. The mode shape changes that occur can be quite large depending on the closeness of the frequencies, the mode shapes, and the magnitude and distribution of the modification (Yang and Griffin [1]). An important example of modal interaction can occur in turbine blades when the blade is “modified” through centrifugal stiffening. In this paper, a multimode and a simpler two-mode model are developed that can be used to predict the effect of centrifugal stiffening on the interaction of bending and torsion modes.

Because turbine blades can behave like cantilevered plates, their lower frequency modes are often referred to as bending or torsion modes. For a “bending” mode, the normal component of the motion is mostly symmetric with respect to the chord centerline whereas the motion of a “torsion” mode is primarily antisymmetric. As engine speed increases, centrifugal stiffening causes a blade’s natural frequencies to increase. Because the frequencies of bending modes tend to increase at a faster rate than those of torsion modes, the frequencies of bending and torsion modes are disposed to cross if the frequency of a bending mode is originally slightly less than the frequency of a torsion mode. Under closer inspection one finds that because of blade asymmetry the frequencies do not cross but veer away from each other without intersecting and in this veering region the modes strongly interact (Kim and Griffin [2]). This phenomenon is known as “frequency veering” or “curve veering” in the literature.

The subject of curve veering has received considerable attention. Perkins and Mote [3] showed that veering can occur in continuous systems and established criteria to distinguish veerings from crossings. Pierre [4] studied the implications of veering on mode localization in periodic systems. Balmes [5] examined the importance of veering and mode localization in the frequency response of damped systems. A publication that provides insight into the interaction of bending and torsion modes in blades is by Nair and Durvasula [6]. They explain frequency crossing and veering in terms of the symmetry group of the mode shapes. According to their argument, three possible behaviors exist. First, if

the two modes are in different symmetry groups, then the natural frequencies of the two modes cross and their mode shapes do not interact. Second, if two modes are in the same symmetry group and the natural frequencies of the two modes come close, their mode shapes interact and their frequencies will veer away from each other. Third, if the structure is not geometrically symmetric, then the mode shapes display neither symmetric nor antisymmetric behavior. In this case, all modes will interact and their frequencies will tend to veer away from each other when they come close. Because of their asymmetry, turbine blades tend to fall into this last category and their modes never cross but always veer away from each other. Typically, one finds that the stronger the asymmetry in the geometry the more pronounced is the veering behavior.

From a practical point of view, frequency veering is only important if it takes place over a broad frequency range. When this happens, the mode shapes change very gradually from bending to torsion and from torsion to bending. As a result, the stress ratios for the modes change with engine speed, more strain gages are necessary for vibration testing, and the strain gage limits used in testing need to change with engine speed (Kenyon [7]).

Finite element analysis is the most practical and common means of calculating the natural frequencies and mode shapes of turbine blades. For systems with large numbers of degrees-of-freedom, finite element analyses can involve long computational times and be fairly expensive. Furthermore, the finite element modal analysis has to be performed at very close intervals of the engine operating speed, especially near the veering region, in order to accurately capture the mode shape changes that occur due to centrifugal stiffening.

In this paper, a multimode model and a simpler, two-mode model are developed which can be used to determine the effect of centrifugal stiffening on the interaction of bending and torsion modes. The model parameters are calculated from a finite element model of the blade. As a result, the frequency veering curves and the modal interaction curves can be calculated for the entire operating range of the engine after only a few finite element calculations.

This paper is organized as follows. The theoretical basis for the multimode and two-mode models is presented in Section 2. Results from the reduced-order models are compared with those directly calculated using the finite element program in Section 3. Conclusions are given in Section 4.

2 Formulation

2.1 Subset of Nominal Modes Formulation. A finite element model of a turbine blade generates stiffness and mass matrices for the blade. In the case of centrifugal stiffening, commercial

Contributed by the International Gas Turbine Institute (IGTI) of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Paper presented at the International Gas Turbine and Aeroengine Congress and Exhibition, Munich, Germany, May 8–11, 2000; Paper 00-GT-611. Manuscript received by IGTI November 1999; final revision received by ASME Headquarters February 2000. Associate Editor: D. Wisler.

finite element codes such as ANSYS or NASTRAN will calculate a stiffnesses matrix that changes with angular acceleration and the change in the stiffness is proportional to the square of the rotational speed, s . As a result, the modes and natural frequencies of the blade satisfy the equation

$$(K^0 + s^2 \Delta K) \cdot \Phi(s) = M^0 \cdot \Phi(s) \cdot [\Omega(s)]^2 \quad (1)$$

where K^0 and M^0 are the stiffness and mass matrices of the blade at zero engine speed and ΔK is the change in the stiffness matrix for a unit change in rotational speed. $\Phi(s)$ and $\Omega(s)$ are the normalized modal and frequency matrices defined as

$$\Phi(s) = [\Phi_1(s) \Phi_2(s) \dots \Phi_N(s)] \quad (2)$$

$$\Omega(s) = \text{diag}[\omega_1(s) \omega_2(s) \dots \omega_N(s)] \quad (3)$$

where N is the number of degrees-of-freedom of the system, and $\Phi_i(s)$ and $\omega_i(s)$ are the i th mode and its associated natural frequency at engine speed, s . It is assumed that the modes are normalized so that all of the modal masses are equal to one. Since the modes at zero engine speed form a complete basis, it is possible to express the modes at engine speed s in terms of these "nominal modes;" refer to Yang and Griffin [8] i.e.,

$$\Phi_j(s) = \sum_{i=1}^N \Phi_i(0) c_{ij}(s) \quad (4)$$

where c_{ij} is the component of the i th mode shape at zero engine speed in the j th mode shape at engine speed, s . Substituting (4) into (1), and using orthogonality by premultiplying by $[\Phi(0)]^T$ yields

$$[\hat{K}^0 + s^2 \Delta \hat{K}] \cdot C(s) = C(s) \cdot [\Omega(s)]^2 \quad (5)$$

where $\hat{K}^0 = [\Omega(0)]^2$, $\Delta \hat{K} = [\Phi(0)]^T \Delta K \Phi(0)$, and $C = (c_{i,j})$ is the modal content matrix. The structural eigenvalue problem is now formulated in terms of the nominal modes of the system.

It was shown in Yang and Griffin [8] that only the nominal modes that have frequencies close to the frequencies of the modes of interest participate significantly in their response. This suggests a computationally efficient procedure for calculating the effect of rotational speed on the natural frequencies and mode shapes of the system that uses only a subset of the nominal modes (SNM). For example, suppose we are interested in the first five modes. We use a finite element code to calculate the change in the stiffness matrix ΔK for a unit change in speed, and the mode shapes and natural frequencies of the first M nominal modes at $s=0$. We use only these first M modes to calculate truncated versions of $\hat{K}^0$ and $\Delta \hat{K}$ that are M by M matrices. We then find the eigenvalues and vectors of the truncated matrices as a function of s . Since M is much less than N this is far more computationally efficient than using the finite element program directly. The results converge to the exact solution as M increases. However, we frequently get good approximations with a very limited number of nominal modes.

Often there are only two modes, a bending and a torsion mode, that interact significantly. In this case, the matrix $\Delta \hat{K}$ can be inferred from frequency information and relatively simple explicit expressions can be developed for the natural frequency and modal content of the two modes. The two-mode model is summarized in the next section.

2.2 The Two-Mode Model. The simplest model that captures the interaction of the torsion and bending mode is one in which we use only two nominal modes to represent the response. As mentioned, the critical situation is when the frequency of the bending mode, ω_1^0 , is slightly lower than the frequency of the torsion mode, ω_2^0 , when the blade is not rotating. In the two-mode model, the matrices in Eq. (5) have the form

$$\hat{K}^0 = \begin{bmatrix} (\omega_1^0)^2 & 0 \\ 0 & (\omega_2^0)^2 \end{bmatrix}, \quad \Delta \hat{K} = \begin{bmatrix} k_{11} & k_{12} \\ k_{12} & k_{22} \end{bmatrix},$$

and

$$C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \quad (6)$$

Our goal is to find C and Ω as functions of the rotational speed, s . The two column vectors in C and the square of the natural frequencies, ω^2 , are the eigenvectors and eigenvalues of the two by two matrix $A(s)$, i.e.,

$$[A(s) - \omega^2 \cdot I] \mathbf{c} = \mathbf{0} \quad (7)$$

where

$$A(s) = \begin{bmatrix} (\omega_1^0)^2 + s^2 k_{11} & s^2 k_{12} \\ s^2 k_{12} & (\omega_2^0)^2 + s^2 k_{22} \end{bmatrix}.$$

Since it is easy to solve the eigenvalue problem for a two-by-two matrix, we can readily calculate the modal content and the natural frequencies of the two modes as a function of the engine speed s once we know the values of k_{ij} and ω_j^0 . The frequencies ω_j^0 ($j=1,2$) can be calculated using any finite element program. Determining the three stiffness terms, k_{11} , k_{22} , and k_{12} is more involved. Specifically, the finite element program has to be able to calculate the effect of blade rotation (centrifugal acceleration) on the stiffness matrix and have an option for writing the stiffness matrix and mode shapes to a file. For example, the commercial finite element program ANSYS has this capability. It is then necessary to write an algorithm that calculates ΔK , the change in the stiffness matrix for a unit change in s , and $k_{ij} = [\Phi_i(0)]^T \Delta K \Phi_j(0)$.

An alternative, simpler approach is to infer the values of k_{11} , k_{22} , and k_{12} from the frequencies ω_1 and ω_2 calculated at two additional engine speeds. Once you have these three parameters you can then calculate the frequencies of the two modes and their modal content for any engine speed by solving (7). The relationship between the stiffness and the frequencies is determined in the following manner.

To find the eigenvalues of (7) we set the determinant of the coefficient matrix equal to zero. This implies that the frequencies satisfy the quadratic equation

$$\omega^4 + b\omega^2 + c = 0 \quad (8)$$

where

$$b = -[(\omega_1^0)^2 + (\omega_2^0)^2 + s^2 \cdot (k_{11} + k_{22})]$$

and

$$c = (\omega_1^0 \omega_2^0)^2 + s^2 \cdot [(\omega_1^0)^2 k_{22} + (\omega_2^0)^2 k_{11}] + s^4 \cdot [k_{11} k_{22} - (k_{12})^2].$$

Solving the quadratic equation gives

$$\omega(s)^2 = \frac{-b(s) \pm \sqrt{D(s)}}{2} \quad (9)$$

where after simplification

$$D(s) = \{[(\omega_1^0)^2 - (\omega_2^0)^2] + s^2(k_{11} - k_{22})\}^2 + 4s^4(k_{12})^2 \quad (10)$$

Also from (9) we can write

$$b(s) = -\{[\omega_1(s)]^2 + [\omega_2(s)]^2\} \quad (11a)$$

and that

$$D(s) = \{[\omega_2(s)]^2 - [\omega_1(s)]^2\}^2. \quad (11b)$$

Consequently, if we use a finite element code to solve for the frequencies of the blade at two different engine speeds, s_1 and s_2 , we can use (11) to find b and D at s_1 and s_2 . Then from (8) and (10) and some algebraic manipulations, we can solve for the stiffness k_{11} , k_{22} , and k_{12} . Let $b_j = b(s_j)$ and $D_j = D(s_j)$ and define

$$E = \frac{-b_1 - [(\omega_1^0)^2 + (\omega_2^0)^2]}{(s_1)^2} \quad (12)$$

$$F = \frac{(s_2)^4 \cdot D_1 - (s_1)^4 \cdot D_2 + [(\omega_1^0)^2 - (\omega_2^0)^2]^2 \cdot [(s_1)^4 - (s_2)^4]}{2 \cdot (s_1 s_2)^2 [(\omega_1^0)^2 - (\omega_2^0)^2] \cdot [(s_2)^2 - (s_1)^2]} \quad (13)$$

$$G = \frac{D_1 - [(\omega_1^0)^2 - (\omega_2^0)^2 + (s_1)^2 \cdot F]^2}{(s_1)^4} \quad (14)$$

then

$$k_{11} = \frac{E + F}{2} \quad (15)$$

$$k_{22} = \frac{E - F}{2} \quad (16)$$

$$k_{12} = \frac{\sqrt{G}}{2} \quad (17)$$

Once the values of the parameters are determined then the frequencies can be calculated as a function of s from (9). Once $\omega(s)$ is known, the modal content is readily calculated from (7) as

$$\frac{c_2}{c_1}(s) = \frac{(\omega(s))^2 - s^2 k_{11} - (\omega_1^0)^2}{s^2 k_{12}} \quad (18)$$

It should be emphasized that the value of the frequency approach for determining k_{ij} in the two-mode model is that it is simple to implement—you only have to run the finite element program at three rotational speeds and note the frequencies of the two modes of interest. It is not necessary to download stiffness matrices and mode shapes nor is it necessary to write special algorithms to compute the matrix products of (5).¹ It also has another important implication. It implies that the mode shape changes that occur in the veering region are only related to the shapes of the frequency curves, i.e., the k_{ij} are determined from the frequencies at three engine speeds and these, in turn, determine the modal content from (18). We will see in the following numerical example that the frequency method of system identification and the two-mode model can, in fact, accurately capture the physical behavior of the example test problem in the veering region.

3 Numerical Study

The geometry and modal properties of the test problem are described in Section 3.1. The results from direct finite element analyses are compared with those from a six-mode SNM analysis in Section 3.2 and with those from a frequency based two-mode model in Section 3.3. Some additional comments on how to generalize the approach to accommodate temperature changes and the effects of coupling the blade to a disk are given in Section 3.4.

3.1 Description of Geometry and Modal Properties. The mode shapes and natural frequencies of a cantilevered tapered plate, Fig. 1, were calculated as a function of rotational speed using the ANSYS finite element program. The geometry of the structure was chosen to be as simple as possible and still provide frequency veering similar to that seen in turbine blades. Specifically, the aspect ratio of the plate (4:1) and its average thickness was chosen so that the second bending frequency, 258 Hz, was slightly less than the first torsion frequency, 333 Hz, when the blade is stationary. The plate was tapered (2:1) in order to make the geometry asymmetric so that the frequencies would veer and

¹For example, if we were to use modal content to determine k_{ij} we would have to download the mass matrix and the mode shapes at zero and some reference rotational speed. Then we would have to compute the modal content by calculating the matrix product indicated in (19). The effort would be comparable to computing k_{ij} directly using (5).

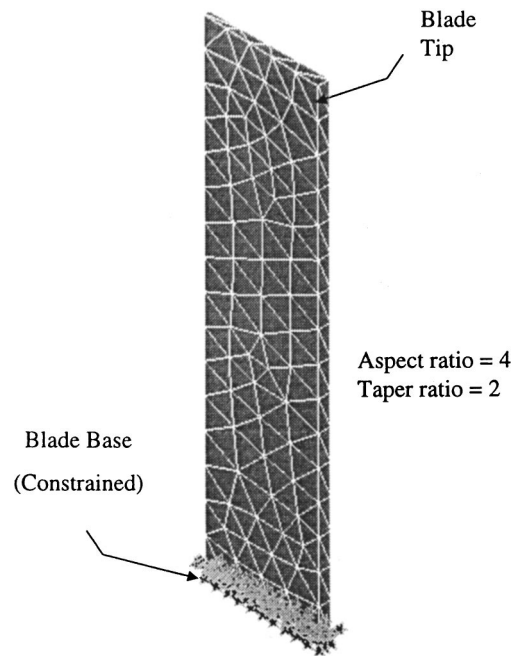


Fig. 1 Geometry of cantilevered tapered plate

not simply cross. To simulate the rotation of the engine the plate is rotated about its constrained base. In order to facilitate the computation, a relatively crude, 2800 degrees-of-freedom, finite element model was used to represent the plate, Fig. 1.

The finite element analysis was used to calculate the frequencies of the first five modes at various rotational speeds, see Fig. 2. From Fig. 2 it is clear that frequencies of second and third modes are relatively close when the blade is stationary and interact as the speed increases to 10,000 rpm. These modes are the focus of our study.

Plots of the frequencies of only the second and third modes are shown in Fig. 3. Contour plots of the mode shapes are superimposed on the graph to indicate how the modes change with speed. From the contour plots we can see that at $s=0$ the lower mode is primarily a second bending mode (symmetric) and the higher mode is primarily first torsion (antisymmetric). At $s = 12,000$ rpm the modes have effectively switched, i.e., the lower frequency mode is torsion and the higher frequency mode is bending. In the veering region, for example, point B, both modes have

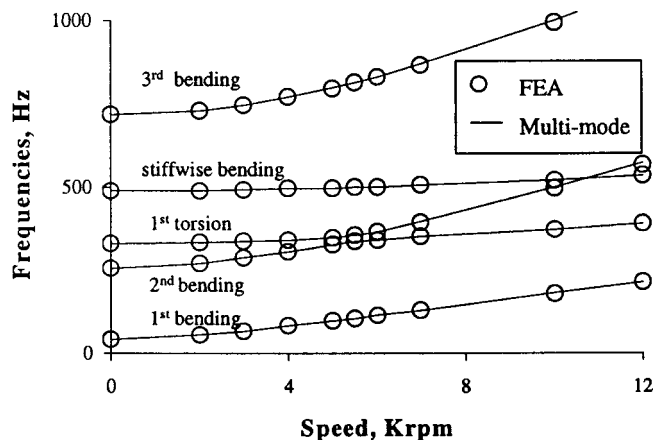


Fig. 2 Frequencies of first five modes versus rotational speed

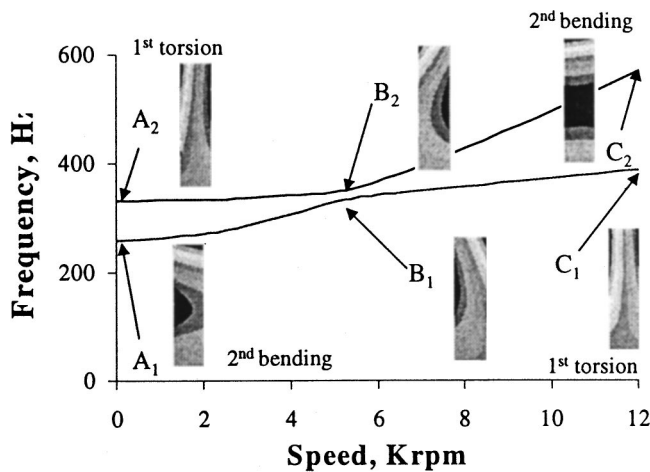


Fig. 3 Frequency veering and mode shape changes

a significant amount of both bending and torsion. It is this frequency veering behavior and the change in the modal content that we want to predict with our simple models.

3.2 Six-Mode SNM Analysis. The finite element program was used to determine the frequencies and mode shapes of the first six modes at $s=0$ and the matrix ΔK . The six-by-six matrices ($M=6$) $\hat{K}_0$ and $\Delta \hat{K}$ were calculated from (5). The resulting algebraic eigenvalue problem (5) was then solved as a function of s using standard MATLAB functions. The computational time required was a few seconds for hundreds of values of s . The computational time required to calculate the natural frequencies in the finite element model can be quite large, depending on the number of degrees-of-freedom.² Typically, the computational time is of the order of tens of minutes per engine speed.

The frequencies from the six-mode model are compared in Fig. 2 with frequencies that were calculated directly from the finite element program. The error in the frequencies for the first three modes was less than 0.1 percent. The maximum error in frequency, less than 1 percent, was observed for the fifth mode. This result is typical for the SNM approach in that modes in the middle of the frequency range are more accurately represented than those at the end of the frequency range, i.e., modes 5 and 6, Yang and Griffin [8]. If more accurate representations of the fifth and sixth modes are needed then M can be increased or the subset of nominal modes that are used can be centered about the mode of interest. For example, the sixth mode would be more accurately represented if nominal modes 4 through 9 were used instead of 1 through 6.

The modal content vectors were also calculated for the six modes and compared with values calculated directly using the centrifugal stiffening option in the finite element analysis. In the case of the direct finite element solution, the modal content was computed using orthogonality, i.e.,

$$c_{i,j} = \Phi_i^T(0) \cdot \mathbf{M}^0 \cdot \Phi_j(s) \quad (19)$$

When the modal content vectors calculated by SNM were normalized to a length of one, the magnitude of the largest error in any component was less than 0.01. Again, the error was larger in the higher frequency modes.

3.3 Two-Mode Model. In the two-mode model developed in Section 2.2 the parameters used in Eq. (7) to determine the frequencies and modal content are determined only from frequency information. In this case the frequencies of the two modes

Table 1 Frequencies of 2nd and 3rd modes

Rotational Speed, s (K-rpm)	2nd Mode (Rad/sec)	3rd Mode (Rad/sec)
0	1621.50	2089.28
3	1803.59	2117.48
5	2066.29	2187.37

at zero and two additional engine speeds must be calculated using the centrifugal stiffening option in the finite element code. This was done for the second and third modes and the resulting frequencies are shown in Table 1.

The modal stiffness parameters k_{ij} were calculated from the values in Table 1 and Eq. (12) through (17). The values of the parameters were $k_{11}=6.980 \cdot 10^4$, $k_{12}=8.314 \cdot 10^3$, and $k_{22}=1.253 \cdot 10^4$ when s is measured in Krpm (thousands of rotations per minute). The frequencies were then calculated from (9) and the modal content from (18). Because of the simple algebraic form of these equations the algorithm was written using MATH-CAD on a PC. It took only seconds to compute and plot the frequencies and modal content as a function of rotational speed. Figure 4 shows a comparison of the frequencies calculated from the two-mode model and those calculated with the finite element analysis. The maximum error in either frequency was 0.35 percent at a rotational speed of 10,000 rpm.

The components of the modal content vectors are plotted as a function of rotational speed in Fig. 5. The circular points were calculated directly from the finite element analysis and Eq. (19). The solid lines connect points that were calculated using the two-mode model and Eq. (18). Note that it appears that there are only three components plotted in the figure when, in fact, there are four. This is because c_{11} is nearly equal to c_{22} , i.e., the amount of bending in the lower frequency mode stays equal to the amount of

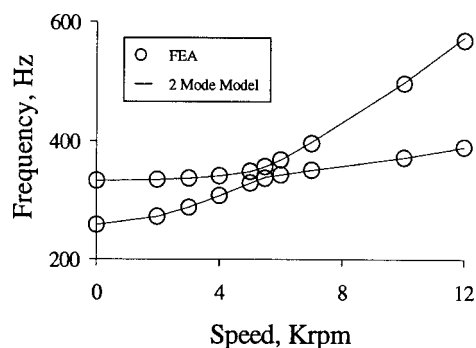


Fig. 4 Comparison of frequencies

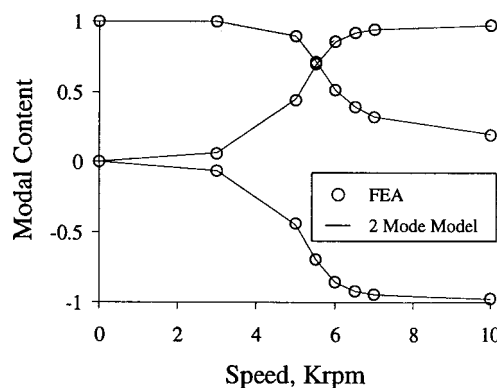


Fig. 5 Comparison of modal content

²The finite element models used in industry are getting progressively larger. For example, we were recently told by a colleague in industry that one of their models of a single blade and disk sector had over 1,000,000 degrees-of-freedom.

torsion in the higher frequency mode as the modes change because of centrifugal stiffening. Because the eigenvectors from the two-mode model are orthogonal this also means that $c_{12} \approx -c_{21}$. It is clear from the plot that the two-mode solution provides a good approximation of the mode shape changes in the test problem. A numerical examination of the data indicates that the results from the two-mode model do, in fact, agree with the finite element calculations to within 1 percent.

3.4 Additional Considerations. The ROM has been shown to accurately represent centrifugal effects over a broad range of operating speeds for the tapered plate test geometry that was studied. We also tested the two-mode ROM against frequency data for a low aspect ratio fan blade that was provided by the U.S. Air Force. The data included pairs of frequencies for the two modes at 0, 20 percent, and 40 percent of maximum speed and these were used to develop the ROM. Strong veering behavior occurred between 75 percent and 100 percent speed. The two-mode model predicted the two frequencies within 0.25 percent over the entire speed range.

It is clear that the ROM captures the effects of centrifugal stiffening quite accurately. What about temperature changes in hot stages? Temperature changes cause the elastic modulus of the material to change and this also has an effect on the frequencies and mode shapes. The simplest method to accommodate temperature effects is to use nominal modes that are calculated for a rotational speed in the middle of the operating range of interest, e.g., $s = s_0$. Then the elastic properties reflect the correct temperatures—at least to first order. This is easily accommodated with the current theory since the only changes in the mathematics of Section 2 are that s^2 actually equals $(s^2 - s_0^2)$ and that it can assume negative values.³

The ROM can be simply extended so that it applies to the modes of a tuned⁴ bladed disk. The approach would be to model a single blade and disk sector and use cyclic symmetric constraints to calculate the natural frequencies and mode shapes, a fairly common option in commercial finite element programs. Then the ROM could be applied to any group of modes that have the same cyclic phase constraint. (Modes with different cyclic constraints do not interact.) The two-mode model can be directly applied since it only requires frequency information. If the multimode ROM is used then the formulation has to be slightly modified to accommodate the complex modes, Yang and Griffin [8]. The key change would be in Eq. (5) where $\Delta \hat{K}$ would equal $[\Phi(0)]^H \Delta K \Phi(0)$ instead of $[\Phi(0)]^T \Delta K \Phi(0)$. H indicates the Hermitian of the matrix.

4 Conclusions

A new, reduced-order model (ROM) has been developed that is based on the subset of nominal modes approach, Yang and Griffin [8]. It can be used in conjunction with standard finite element codes to quickly calculate the effect of centrifugal stiffening on the frequencies and mode shapes of the blade.

The information required to develop the reduced-order model depends on the number of modes used in the representation. Typi-

cally, modes tend to interact in pairs over limited speed ranges and a two-mode ROM can accurately capture this interaction. When a two-mode ROM is used, the parameters can be calculated simply from the frequencies of the two modes determined from a finite element modal analysis at three different rotational speeds. Once the model is determined then the frequencies of the two modes and the modal content can be quickly calculated over a wide range of speeds. The expression for the modal content can be used to give the modal stresses in the blade as a function of engine speed, especially in veering regions where they rapidly change. This approach could be used to determine strain gage limits that more realistically reflect blade modal stresses that vary with the engine's speed. The fact that the two-mode model can be completely inferred from only frequency information has the interesting implication that the change in the modal content is inherently determined by the shape of the frequency veering curves.

To develop an ROM with more than two modes requires that the stiffness matrix in the finite element program be accessed and used to calculate ΔK in (1), i.e., the change in the stiffness matrix for a unit change in rotational speed. This information is used with the mode shapes of the stationary blade to form an algebraic eigenvalue problem that includes the effect of centrifugal stiffening. The number of degrees-of-freedom in the ROM is equal to the number of modes. In an example this procedure was used to reduce a finite element model with 2800 degrees-of-freedom to a ROM with six degrees-of-freedom. The resulting ROM was used to calculate the frequencies and modal content of the first six modes over a speed range of 10,000 rpm. The maximum errors in the frequencies and the modal content of the six modes were less than 1 percent.

In conclusion, a reduced-order model has been developed that can be used to accurately and quickly calculate the changes in the natural frequencies and mode shapes of a blade that are caused by centrifugal stiffening. It has been corroborated by comparisons with finite element analyses of a cantilevered tapered plate and with frequencies from a low aspect ratio fan blade.

Acknowledgments

This work was supported by the U.S. Air Force. The authors wish to thank Major G. Minkiewicz for bringing this topic to their attention and for his encouragement and guidance.

References

- [1] Yang, M.-T., and Griffin, J. H., 1997, "A Normalized Modal Eigenvalue Approach for Resolving Modal Interaction," *ASME J. Eng. Gas Turbines Power*, **119**, pp. 647–650.
- [2] Kim, N., and Griffin, J. H., 1994, "Sensitivity of Bonded and Composite Beams," *J. Sound Vib.*, **177**, pp. 71–92.
- [3] Perkins, N. C., and Mote, C. D., 1985, "Comments on Curve Veering in Eigenvalue Problems," *J. Sound Vib.*, **106**, pp. 451–463.
- [4] Pierre, C., 1988, "Mode Localization and Eigenvalue Loci Veering Phenomena in Disordered Structures," *J. Sound Vib.*, **126**, pp. 485–502.
- [5] Balmes, E., 1992, "High Modal Density Curve Veering, Localization: A Different Perspective on the Structural Response," *J. Sound Vib.*, **161**, pp. 358–363.
- [6] Nair, P. S., and Durvasula, S., 1972, "On Quasi-Degeneracies in Plate Vibration Problems," *Journal of Mechanical Science*, **15**, pp. 987–992.
- [7] Kenyon, J., 1999, "Investigation of Curve Veering Using Computational and Experimental Techniques," *Proceedings of the 40th AIAA/ASME/ASCE/AHS Structures, Structural Dynamics and Materials Conference and Exhibit*, AIAA, Apr., pp. 1550–1558.
- [8] Yang, M.-T., and Griffin, J. H., 1999, "A Reduced Order Model of Mistuning Using a Subset of Nominal System Modes," *ASME Paper No. 99-GT-288*.

³This is not a problem algebraically since s always appears as squared in all of the equations.

⁴The term "tuned" is used to indicate that all blades are identical.

A Gas Turbine-Based Hybrid Vehicle—Part II: Technological and Configuration Issues

R. Capata

e-mail: roberto.capata@uniroma1.it

E. Cioffarelli

E. Sciubba

Department of Mechanics and Aeronautics,
University of Roma 1—"La Sapienza,"
Via Eudossiana, 18,
00184 Roma, Italy

A hybrid propulsion system for a passenger car is described and its preliminary design and performance discussed. A gas turbine operating in a controlled on-off mode, a generator, a battery pack, an inverter, and an electric motor constitute the power system, with an electronic vehicle management unit supervising the system for control and regulation. Road simulation tests are presented, based on standard ECC driving mission cycles. The turbogas performance might be improved adopting ceramic blades, with a possible increase of the TIT (turbine inlet temperature), and a consequent increase of its efficiency. The paper reports on the first results of a study in progress, within a joint research program between the University of Roma 1 "La Sapienza" and the Italian Research Centre of ENEA-Casaccia. [DOI: 10.1115/1.1584475]

1 Introduction

Electric vehicles enjoy some important advantages with respect to traditional ones, such as noise and pollution reduction, kinetic-to-electric braking energy recovery, low maintenance, and relatively high performance. More efficient and much cheaper battery packs with higher energy densities will be needed for commercialization of such vehicles though. Currently available batteries do not show real market competitiveness. Hybrid vehicles ("HV") may be considered an attractive medium term answer to meet environmental and energy consumption, and some studies, [1], contend that the adoption of such vehicles could reduce transportation-related pollution. HVs offer greater power control flexibility, due to their intrinsic ability to activate, simultaneously or separately, different kinds of engines. However, they currently suffer from low system reliability and high maintenance requirements, especially as far as the batteries storage system is concerned. They are also hampered by very high initial commercialization costs.

A hybrid vehicle can be designed for pure city use or for more flexible operation: This is because one can choose among some different hybrid propulsion system configurations, depending on the type of vehicle, on its intended operational mode, and on the executable mission. Configurations range thus from dual to series, parallel and parallel with thermal engine working at fixed point, [2]. At present the most common design consists either of a small displacement engine that supplies a suitable power in a city drive with low pollution emission but offers inadequate performance in highway traffic, or of a larger engine which provides satisfactory highway driveability but is less economic in urban usage, with higher emissions and a lower fuel economy.

2 The Gas Turbine-Based Series Hybrid Propulsion System

At present, reciprocating internal combustion engines dominate the automotive market, thanks to their flexibility of operation and high specific output (approximately 1 kW/kg). However, they are hindered by some important drawbacks, like a high number of accessories and auxiliaries, a high fuel consumption, and relatively high polluting emissions.

One can construct different feasible hybrid configurations by considering different combinations of thermal and electric engines, [2,3], (Fig. 1). In a series hybrid, a thermal engine drives an electric generator that produces electricity to recharge the battery package and/or to power the electric motor.

In the parallel hybrid the wheels, if necessary, can get additional power directly from the thermal engine, while in a dual power hybrid the two thermal and electric propulsion systems are not mechanically coupled.

An interesting solution is the coupling of a turbogas set (TG) with a high-speed turbo-generator, in a series configuration, [2], (Fig. 2). A vehicle powered by such a TG-hybrid system uses a purely electric traction system in both urban and highway missions, and the TG is switched on only when strictly necessary, to recharge the battery pack and to provide the generator with the right requested power, if the battery capacity is not sufficient. Thus, the TG can be designed to operate almost at fixed point most of the time. The characteristics of TG sets that make them attractive for hybrid traction are their mechanical simplicity, their high power density (power-to-volume ratio), reduced low frequency noise and vibrations levels, high specific power (power-to-weight ratio), relatively high efficiency, the absence of a water cooling system, the practical absence of lubrication (limited to shaft bearings), the stationary combustion, and their multi-fuel capability. On the other side TG sets display some disadvantages, including a relatively low off-design efficiency, a high rotational and thermal inertia, extremely low braking capacity, safety problems, and (for the low power outputs considered here) a high initial cost.

The scheme of the hybrid system proposed in this series of papers is the following: the vehicle management unit turns the TG on or off, and controls the percentage of the turbogas mechanical power sent directly to the electric motor and that used to recharge the batteries. In the following we assume that at any instant the TG is able to supply the total power requested by both the electric motor and the battery pack, resulting in the right power available at the right time.

3 Results of the Simulations

The results discussed below have been obtained by means of a simulator in C++ that reproduces the instantaneous energy flows in the system to calculate the power demand at the wheels, depending on the speed and acceleration demands imposed by standard urban and extra urban driving cycles, [2], (Table 1, Fig. 3). Simulations include energy balances and a preliminary turbogenerator sizing. These calculations take just a few seconds of run time on a Pentium III 1600 MHz PC unit.

¹To whom correspondence should be addressed.

Contributed by the International Gas Turbine Institute of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by IGTI Nov. 2001; final revision received by the ASME Headquarters Dec 2002. Associate Editor: R. Gabriellson.

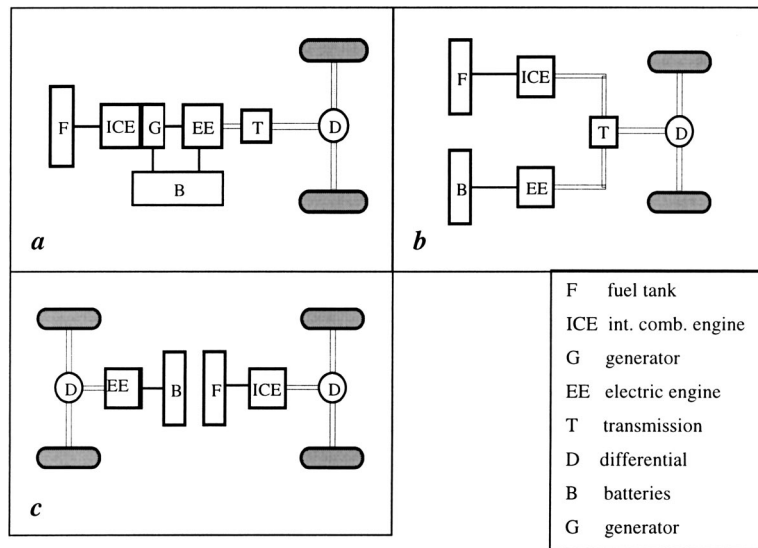


Fig. 1 (a) Series hybrid system, (b) parallel hybrid system, (c) dual power hybrid system

At each instant in time, the required power in each component is obtained by a dynamic analysis of the vehicle that takes in account aerodynamic drag, tire rolling resistance and vehicle mass, [4]. A continuous TG regulation [2] seems necessary to

maintain a high TG efficiency even in case of low total power demand: such a regulation could be obtained with fuel flow control or/and with a variable geometry turbine and IGV blades at compressor inlet, [5]. We present here only the outcome of the numerical simulations in TG on-off mode (for details see [2]): the vehicle starts its mission in a purely electric mode, with fully charged lead-acid batteries. The TG is switched on if the state of charge (SOC) falls below 60%, and is switched off again when the batteries recover to a SOC of 90%.

Several simulation runs were completed to correctly assess the performance of the proposed propulsion system. Two basic mission profiles were analyzed: ten consecutive urban cycles (ECE15) and a “complex driving mission” composed of four consecutive urban cycles (ECE15), followed by four consecutive extra-urban cycles (EUDC) and then by 72 minutes of continuous highway drive at 120 km/hr. After a series of trials, a 25 kW TG was chosen, coupled with a 200 kg battery pack, that delivers 30 kW. This is a good compromise between the needed battery weight and TG power. The results of these simulations are reported in Figs. 4 and 5, with auxiliary tables providing all necessary information to evaluate the TG performance.

The important result of the concurrent TG and batteries power optimization is that the TG must not be switched on and off frequently during urban cycles, and that it remains always “on” during extra urban and highway drives, with batteries excluded by a manual control or by the vehicle management unit. For instance, power flow from the battery pack to the motor could be turned off when vehicle speed remains above a certain value for a certain time, to be turned on again when vehicle speed stays below a lower limit for a longer period of time. Such an “intelligent” control is definitely within the reach of current electronic logic systems. If the design goal is the construction of a city car, i.e., of a vehicle not capable of substantial speed and acceleration performance, thermal and electric power outputs can be lower: a 10 kW TG/80 kg battery pack configuration was tested for the design vehicle specifications given in Table 1, and a 10-ECE15 cycles-simulation demonstrated that the vehicle performs well, in an urban cycle, as shown in Fig. 6. Obviously, its performance in an extended extra-urban mission is marginal.

The interesting results are those related to fuel economy: Table 2 shows a comparison between the results of a series of simulations with mixed cycles (36.8% urban and 63.2% extra urban) in the TG-HV configuration proposed here. Both 1000 kg and 1400 kg vehicles were simulated, and the results are compared with some published commercial data.

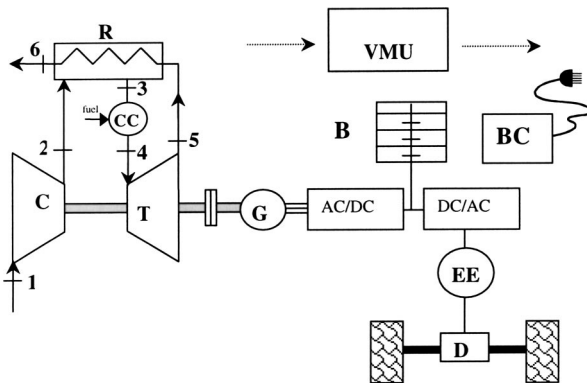


Fig. 2 Hybrid gas turbine vehicle scheme. C=compressor, T=turbine, R=regenerator, CC=combustion chamber, G=generator, AC/DC inverter, B=batteries, BC=battery charger, EE=electric engine, D=differential, VMU vehicle management unit.

Table 1 Design specifications

Wheel rolling radius	$r = 0.265 \text{ m}$
Vehicle width	$b = 1.7 \text{ m}$
Vehicle height	$h = 1.4 \text{ m}$
Net front area	$S_f = 2.142 \text{ m}^2$
Area ratio (S/S_{tot})	$\alpha = 0.9$
Aerodynamic drag	$c_x = 0.25$
Tire rolling friction coefficient	$f_r = 0.015$
Vehicle mass	1000–1400
Equivalent mass	$m_e = 1600 \text{ kg}$
Maximum power demand	P_{max}
TG power	$P_{TG} = P_{max}$
Battery power	$P_{bt} = P_{max}$
Module weight	$m_{mod} = 20 \text{ kg}$
Module voltage	$V_{mod} = 12 \text{ V}$
Cell voltage	$V_{cel} = 2 \text{ V}$
Specific power	$P_{sp} = 150 \text{ W/kg}$
Specific energy	$E_{sp} = 30 \text{ Wh/kg}$
TG switch-on setting	SOC < 60%
TG switch-off setting	SOC > 90%

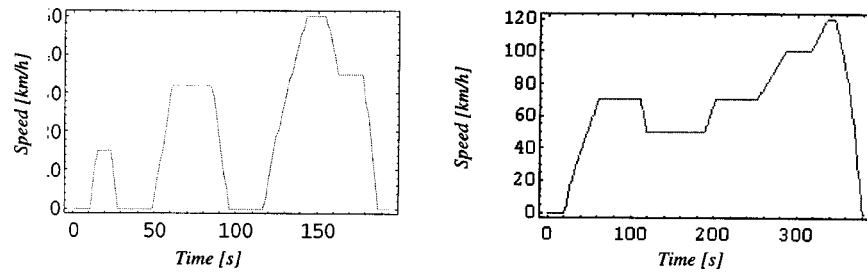


Fig. 3 ECC urban and extra-urban cycles

Cycle efficiency is improved if the compressor is staged and intercooled, as shown in Fig. 6.

An additional improvement in cycle efficiency can be also achieved by inserting a regenerator to preheat the combustion air: Figure 7 shows the process details.

There are two types of regenerators: static, usually with a cross-flow arrangement, and rotating ones, more expensive but with a lower weight-to-power ratio. It is likely that rotating regenerators would be the most convenient choice for vehicular application: but only a thermo-economic comparison between all possible configurations can provide all necessary information to support one choice over the other.

It is obvious that the introduction of such devices means a constructive complication (larger size, higher installation cost), but brings an advantage in terms of specific consumption: thus, we are faced again with a problem of thermo-economic optimization.

A model for the calculation of emissions has also been developed. The code ("TGEEC") simulates the steady-state combustion process in a TG, and supplies step by step the mixture characteristics (temperature and molar fraction of each species) during the chemical process. The model includes 14 equations for methane combustion, while thermal NO_x are modeled after Zeldovitch, [1]. No further configuration analysis has been performed: It has been assumed that the CH_4 tank can be fitted into the vehicle without major modifications (Table 3).

4 Ceramic Gas Turbines

Demonstration level tests are currently being run under several research programs, both in Japan and in Europe, [6–8], with the

goal of assessing possible mechanical and thermal improvements due to the adoption of ceramic components, for rotors, nozzles, and combustion chamber.

High temperature strength and endurance properties tests confirm the possibility of a substantial increase of turbine inlet temperature (TIT), to 1623 K, with operating rotor speeds equal to those of current metallic rotors: some materials such as SN88M and Si_3N_4 , have proven to be quite suitable for this purpose.

Such a high TIT obviously increases TG efficiency, with clear advantages in terms of fuel consumption and pollution, as shown in Figs. 8 and 9.

5 Conclusions

The technical feasibility of a TG based hybrid propulsor for a passenger car has been demonstrated and its performance optimized. In all cases examined here, the average power provided by the TG was different from the maximum one, and this indicates the necessity of a power modulating system. The TG could work at its optimal fixed point only for a constant low power level, which implies a very large battery pack to satisfy the vehicle power demand.

All simulations indicate a significant fuel economy and a substantial reduction of the vehicle emission levels.

The steps of the Hybrid Vehicle Project launched by ENEA and the University of Roma 1 are the following:

1. Perform a comparative thermodynamic analysis of several configurations and select the most promising one(s).
2. Perform a feasibility analysis of the selected configurations.
3. Perform a complete mission analysis to establish perfor-

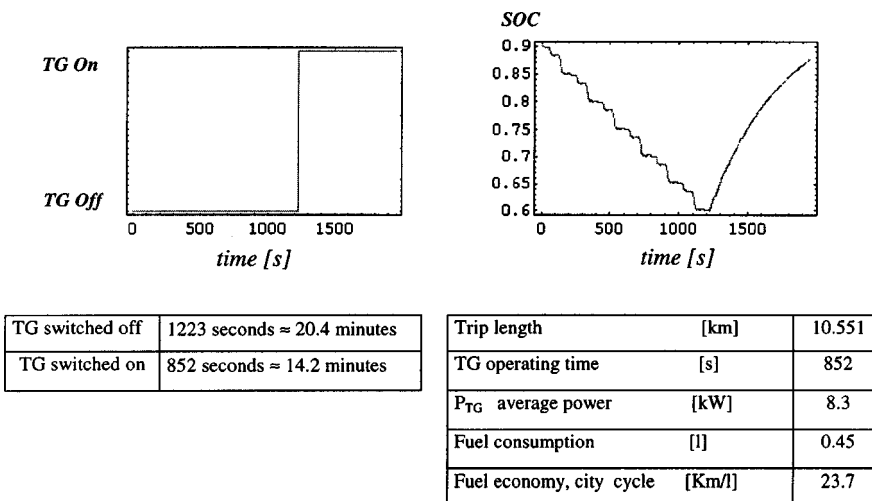


Fig. 4 10 urban cycle: $P_T=25$ kW, $m_{bt}=200$ kg

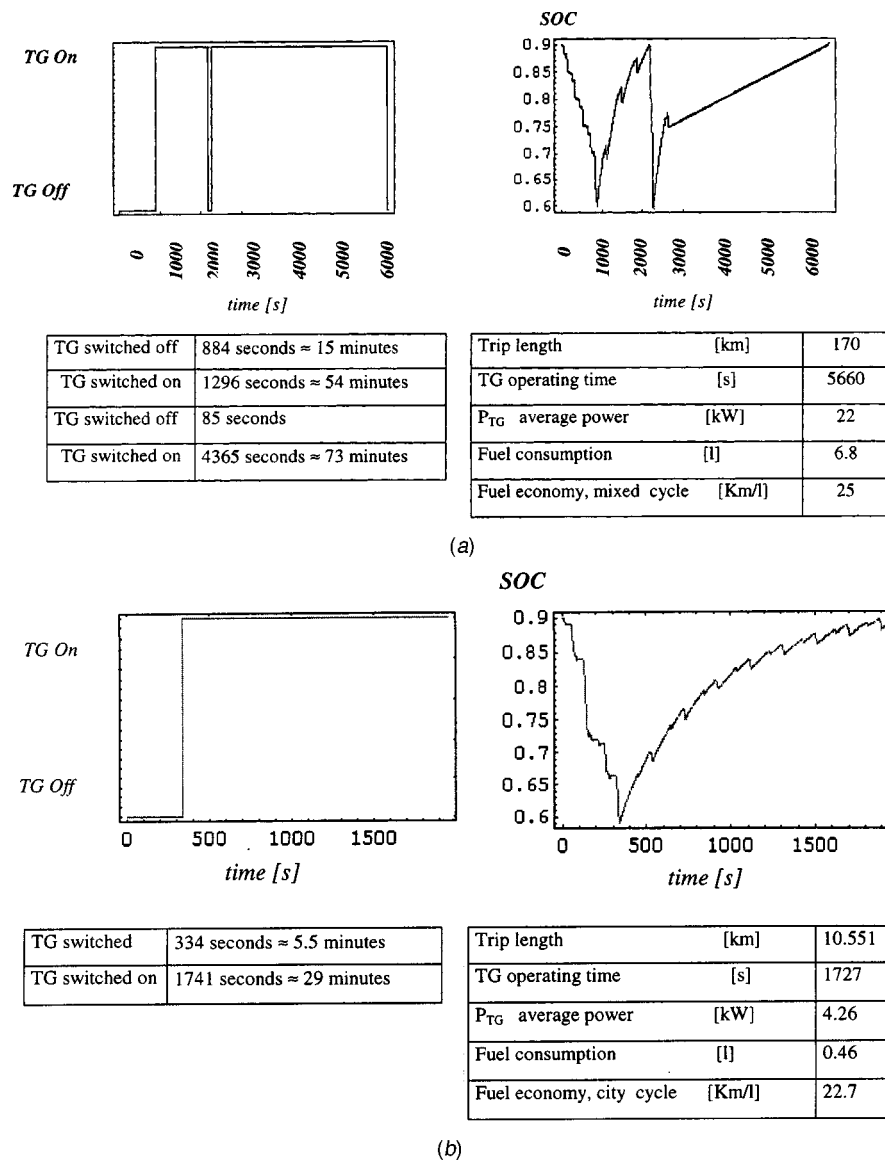


Fig. 5 (a) Mixed urban highway cycle: $P_T=25$ kW, $m_{bt}=200$ kg, (b) 10 urban cycles: $P_T=10$ kW, $m_{bt}=80$ kg

mance. Finalize a small number of configurations (two or three).

4. Perform a thermo-economic analysis of each surviving configuration to assess its economic cost/benefit ratio.
5. Build and test one or two configurations. The testing is to be carried out at several levels: first, the TG (on a test-rig) and

the batteries alone, then the entire system at design point, then a limited number of off-designs points, then the entire system mounted on a test vehicle.

This paper has described Step 3 above. Steps 1 and 2 have been

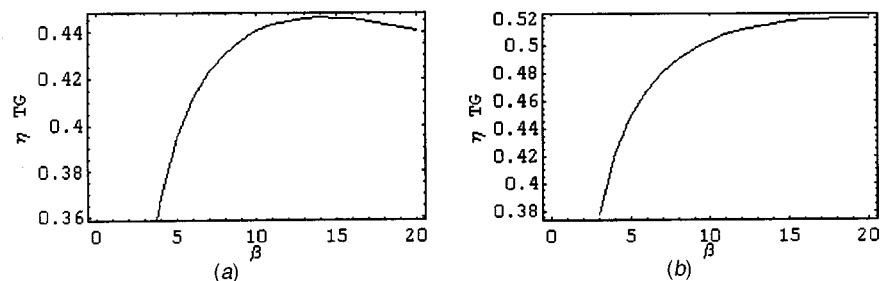


Fig. 6 Turbogas efficiency curve. $P_{el}=25$ kW, $n=100,000$ rpm. (a) Without IC, (b) with IC.

Table 2 Mixed cycle comparison among some commercial vehicles and the TG hybrid vehicle

Vehicle	Engine	Mass kg	Fuel Economy km/l
Audi A4	2.5 TDI	1495	13
Ford Galaxy	1.9 TD	1650	15.2
Seat Ibiza	1.9 TDI	1124	20.0
Fiat Punto	1.9 JTD	1040	20.4
Ford Focus	1.8 TD	1183	20.4
Smart	600 TI	720	20.8
VW Lupo	1.4 TDI	947	23.3
Toyota Prius	1500 cc 43 kW, 30 Kw batteries	1262	23
Volvo Power Split	1400 cc 56 kW, 60 Kw batteries	N.A.	N.A.
Honda Insight	1000 cc 63 kW, 10 Kw batteries	860	29
TGHV	25 KW TG, 30 kW batteries	1000	36.2
TGHV	25 KW TG, 30 kW batteries	1400	29.2

*N.A. not available

reported in a previous paper, [3]. At present, we are carrying out the two remaining steps, which ought to be completed within two years.

It is still unclear whether the TG hybrid vehicle can be a serious midrange competitor for the ICE., but preliminary results and the

Table 3 TG design specifications

Rotational speed, n	100000 rpm
Nameplate power, P	25 kW
Maximum allowable peripheral velocity, $U_{\max}$	500 m/s
Fuel lower heating value, LHV	50241 kJ/kg
Inlet pressure, p_1	1 bar
Compressor delivery pressure, p_2	4.4 bar
Inlet temperature, T_1	288 K
Maximum allowable T_4	1400 K
Degree of regeneration, R	0.5
Maximum allowable shaft torsional stress, τ	40 MPa
Turbine polytropic efficiency, η_{pc}	0.9
Compressor polytropic efficiency, η_{pc}	0.9
Mechanical efficiency, η_m	0.85

scant data available in the recent literature make it clear that at present the coupling TG-electrical batteries-electrical motor is at an advantage with respect to the coupling fuel cell-electrical motor, both for performance and for industrial feasibility. Its practical realizability is actually undisputed, and its real costs can only be assessed after certain number of prototypes have accumulated enough mileage on the road. Its chances of success probably de-

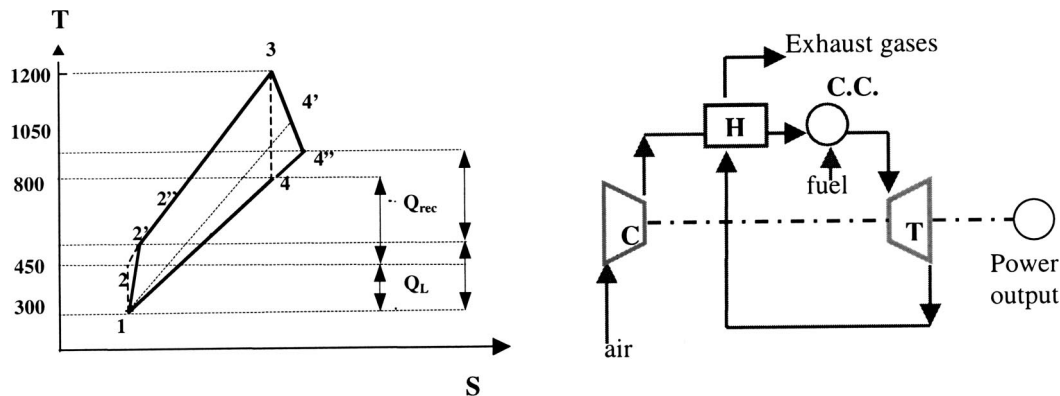


Fig. 7 TG cycle

TIT K	m_{comb} g/s	Fuel consumption g/kWh	urban drive km/l	η_{TG}
1423	1.58	228	20.8	0.369
1523	1.55	224	21.2	0.375
1623	1.53	220	22	0.380

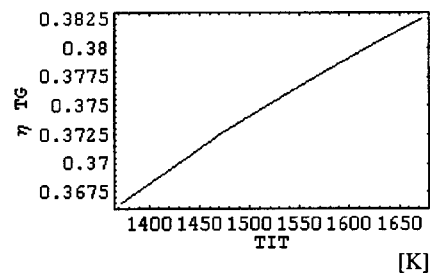


Fig. 8 TG efficiency and fuel consumption without intercooler. PT=25 kW.

TIT K	m_{comb} g/s	Fuel consumption g/kWh	urban drive km/l	η_{TG}
1423	1.18	170	27	0.491
1523	1.17	169	28.5	0.495
1623	1.16	168	30	0.499

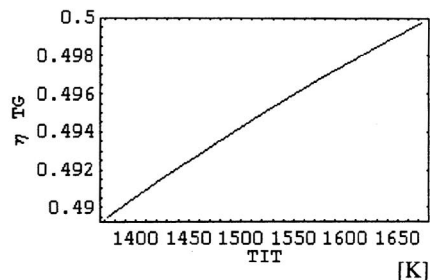


Fig. 9 TG efficiency and fuel consumption with intercooler. PT=25 kW.

depends only on industry long-term strategy (what retooling to select) and on policymaking decisions (how to incentivate a gradual conversion of the existing fleet).

An increase of the TIT to 1623 K, thanks to ceramic components, would improve the gas turbine efficiency, with a consequent fuel consumption and pollution reduction. Prior to launching a pilot project, it was decided to perform a thermo-economic analysis to assess technical and economic features of the different schemes studied in this paper.

References

- [1] Turns, S. R., 1996, *An Introduction to Combustion*, McGraw-Hill, New York.
- [2] Cioffarelli, E., 2000, "Preliminary Design of a Turbogas Based Series Hybrid Vehicle Propulsion System," M.E. thesis, Department of Mechanics & Aeronautics Engineering University of Rome 1 (in Italian).
- [3] Cioffarelli, E., and Sciubba, E., 2000, "A New Type of Gas Turbine Based-Hybrid Propulsion System. Part 1: Concept Development, Definition of Mission Parameters and Preliminary Sizing," Proceedings AES/ASME Winter Meeting, Orlando, FL.
- [4] Hucho, W. H., and Sovran, G., 1993, "Aerodynamics of Road Vehicles," *Annu. Rev. Fluid Mech.*
- [5] Rautenberg, M., and Malobabic, M., 1985, "On Turbochargers With Variable Geometry," ASME Paper No. 85-GT-149.
- [6] Gabrielsson, R., and Holmqvist, G., 1998, "Progress on the European Gas Turbine Program AGATA," *ASME J. Eng. Gas Turbines Power*, **120**.
- [7] Nakazawa, N. et al., 1996, "Radial Turbine in Automotive Development for the 100 kW Automotive Ceramic Gas Turbine," *ASME J. Eng. Gas Turbines Power*, **120**.
- [8] Rautenberg, M. et al., 1984, "Improvement in Compressor Performance of a Small Turbocharger by Using Ceramic Insulation Between Turbine and Compressor," Fifth International Conference for Mechanics Power Engineering, Cairo, Oct.
- [9] Berenyi, S. G. et al., 1998, "Hybrid Vehicle Turbine Engine Technology Support," ASME Paper No. 98-GT-451.
- [10] Willis, G., and Radtke, R., 1985, "Hybrid Vehicle Systems Analysis," SAE Paper No. 850225.

Z. Huang

School of Energy and Power Engineering,
Xi'an Jiaotong University,
Xi'an 710049, P. R. China
e-mail: zhhuang@mail.xjtu.edu.cn

S. Shiga¹

Department of Mechanical Engineering,
Faculty of Engineering,
Gunma University,
Kiryu, Gunma 376-8515, Japan
e-mail: shiga@me.gunma-u.ac.jp

T. Ueda

General Research Center,
Nissan Motor Co., Ltd.,
Yokohama-city,
Kanagawa 230-0053, Japan
e-mail: ta-ueda@mail.nissan.co.jp

H. Nakamura

e-mail: nakamura@me.gunma-u.ac.jp

T. Ishima

e-mail: ishima@me.gunma-u.ac.jp

T. Obokata

e-mail: obokata@me.gunma-u.ac.jp

Department of Mechanical Engineering,
Faculty of Engineering,
Gunma University,
Kiryu, Gunma 376-8515, Japan

M. Tsue

e-mail: tsuem@mail.ecc.u-tokyo.ac.jp

M. Kono

e-mail: tmkono@mail.ecc.u-tokyo.ac.jp

Department of Aeronautics and
Astronautics Engineering,
The University of Tokyo,
Tokyo 113-8656, Japan

Effect of Fuel Injection Timing Relative to Ignition Timing on the Natural-Gas Direct-Injection Combustion

The effect of fuel injection timing relative to ignition timing on natural gas direct-injection combustion was studied by using a rapid compression machine (RCM). The ignition timing was fixed at 80 ms after the compression start. When the injection timing was relatively early (injection start at 60 ms), the heat release pattern showed a slower burn in the initial stage and a faster burn in the late stage, which is similar to that of flame propagation of a premixed gas. In contrast to this, when the injection timing was relatively late (injection start at 75 ms), the heat release rate showed a faster burn in the initial stage and a slower burn in the late stage, which is similar to that of diesel combustion. The shortest duration was realized at the injection end timing of 80 ms (the same timing as the ignition timing) over a wide range of equivalence ratio. The degree of charge stratification and the intensity of turbulence generated by the fuel jet are considered to cause this behavior. Early injection leads to longer duration of the initial combustion, whereas late injection leads to a longer duration of the late combustion. Early injection showed relatively lower CO concentration in the combustion products while late injection gave relatively lower NO_x. It was suggested that early injection leads to combustion with weaker stratification, and late injection leads to combustion with stronger stratification. Combustion efficiency was kept at a high value over a wide range of equivalence ratio. [DOI: 10.1115/1.1563243]

Introduction

With stringent legislation of automobile pollutants and the need for continuous improvement of thermal efficiency of internal combustion engines, more comprehensive and detailed research work has to be conducted. Diesel engines have the advantage of high thermal efficiency; however, the problems of NO_x emissions and particulate matter still remain. Traditional premixed charge combustion engines have lower thermal efficiency due to the occurrence of engine knock and the unavoidable throttling at the intake. In recent years, so-called homogeneous lean combustion engines have been developed, [1–4]. These engines can attain a higher thermal efficiency due to the lower pumping loss, lower heat loss, and an increased ratio of specific heat, at the expense of moder-

ately high NO_x emissions due to the ineffectiveness of the catalyst. However, the lean operation limit of this type of lean homogeneous engines was not really high compared to that of diesel engines. In the past few years, the direct-injection gasoline engine has entered into the production stage for both automobile and two-stroke marine engines. The charge stratification in the combustion chamber permits extremely lean combustion which increases the thermal efficiency, but gives the penalty of NO_x and even particulate emissions, [5–9]. Therefore much work still needs to be done to address this emission problem.

Natural gas is regarded as one of the most promising alternative fuels and is also considered one of the cleanest fuels. Furthermore, its high-octane value permits a high compression ratio leading to higher thermal efficiency at full load conditions. In previous studies even lower emissions could be realized with natural gas, [10,11]. Nowadays, two types of natural gas engines are primarily studied: One is a dual-fuel diesel engine using homogeneous natural gas mixture and diesel fuel spray, [12,13], and the other is a homogeneous spark-ignition engine, [1,10]. The former needs two fuel supplying systems and still has a problem of high HC emis-

¹To whom correspondence should be addressed.

Contributed by the Internal Combustion Engine Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the ICE Division October 2001; final revision received by the ASME Headquarters September 2002. Associate Editor: D. Assanis.

sions; the latter has problems of low power output and combustion instability under lean operating conditions which are basically the same as that of gasoline lean-burn operation.

Many studies on diesel combustion have revealed its dual combustion pattern of premixed phase and diffusion phase. These studies have shown that in ordinary diesel engines NO_x increases with advancing the injection timing due to the increase in the fuel burned in premixed phase, while CO and smoke increase with retarding the injection timing due to less mixing during the diffusion phase, [14]. However, when advancing the fuel injection timing much earlier than ordinary, like in the case of HCCI combustion, the combustion becomes uncontrollable and nearly homogeneous with little NO_x and smoke, [15], and thus a different combustion pattern from that of classical diesel combustion is realized. Consequently, the injection-timing change can generate not only a shift of these two kinds of combustion patterns in direct-injection diesel engines, but also possibly a different pattern of combustion. Thus the injection timing effect on combustion is complex even in diesel combustion, which means that it is difficult to simply state that earlier injection tends to produce higher NO_x and late injection tends to increase CO and smoke.

In the case of direct-injection gasoline spark-ignition combustion, however, only a few studies dealing with basic combustion behavior can be found. The authors have done a study in this category using a test engine, [16], but studies which systematically reveal the effect of injection-timing on combustion are hard to find. Furthermore, few basic studies in the field of spark-ignition direct-injection gasoline engine utilizing a combustion bomb or a rapid compression machine (RCM) can be found. Many aspects of this phenomenon still have not still been clarified and thus need to be investigated. This is presumably due to a very complicated phenomenon with one parameter, i.e., spark timing, more than that of diesel combustion. Thus the authors consider that there is much need for basic combustion research in the field of gasoline direct-injection spark-ignition engines.

Although there are a few studies on direct-injection gas engines, more knowledge or information on the basic combustion characteristics is necessary. Agarwal and Assanis [17] developed a model of natural gas ignition, combustion, and pollutant formation in direct-injection engines, while Goto [18] studied the stability of combustion versus fuel injection timing in a direct-injection natural gas engine. These two studies gave little information in terms of the injection timing effect. Recently in the field of basic studies of natural gas combustion utilizing an RCM, Brett et al. [19] has made an improved chemical kinetic model to simulate natural gas spontaneous ignition and compared the RCM experimental results obtained under a wide range of temperature and pressure conditions made with various composition of methane, oxygen, and argon. Although this study does not give direct information for understanding stratified-combustion, it will serve as a first step toward future work on knock occurrence conditions.

There is a great difference in fuel properties and possibly in mixture preparation processes between gaseous and liquid fuels. Furthermore, almost no basic study can be found on the effect of injection timing of gaseous fuel. Thus, both theoretical analysis and experimental investigation of the effect of injection timing are required to enhance our understanding of the direct-injection CNG spark-ignition stratified-combustion.

A preliminary study has been carried out by the authors, [20], and several interesting results were obtained. One of them is that an extremely lean combustion was attained, and the lean limit was just $\phi = 0.1$ with satisfactory combustion efficiency greater than 0.92. However, since those studies were just at the stage of feasibility, many aspects worth examining remain. One of them is the fuel injection timing relative to ignition timing on combustion in direct-injection natural gas stratified combustion. Thus, the objective of the present study is to reveal the effect of the fuel injection timing relative to ignition timing on the combustion characteristics in natural gas direct-injection combustion using an RCM.

Experimental Setup and Procedures

Since the details of the experimental procedures are presented in the previous study, [20], only a brief description is given here.

RCM Specifications, Injection System, and Ignition System.

Specifications of the rapid compression machine used in this study are given in Table 1 and the top view of the combustion chamber is shown in Fig. 1.

A spark gap was located at the center of the combustion chamber and only the opposed injection (O. Inj.) mode with two injectors arranged in opposite was used. The injector used in this study is a modified version from a gasoline direct-injection engine. To increase the flow rate for CNG application, the swirler near the tip was taken off. To drive the injector, a pulse with TTL level was input into the injector driver which works with normal 12 V DC and does not require any voltage enhancement. The calibration of the pulse width with the injection amount was not only made by the manufacturer (Hitachi Co.) but also by our laboratory, since the amount of fuel affects so many parameters with high sensitivity. The fuel was injected into a vessel with a volume of 34.98 L, and the amount of injected natural gas was determined by the pressure increase of the vessel content as well as the gas analysis with a gas chromatograph after mixing well. Thus the relationship between the injection amount and the pulse width was obtained, and the results showed an almost linear relationship between the injection amount and the pulse width over the equivalence ratio of 0.1 to 1.0. When the equivalence ratio was smaller than 0.1, slight deterioration from this linearity was observed.

The ignition system used was the same as that of a commercial gasoline direct-injection engine without any modification. It was a normal full transistor type. An inductive spark was supplied with a sufficient dwell duration of 12 ms. This was expected to give the spark energy comparable to a normal gasoline engine. The spark gap for this RCM experiment was set at 0.5 mm, which was determined by the level of noise and possibility of break down outside of the combustion chamber.

Table 1 Specifications of rapid compression machine

Cylinder bore/mm	80
Stroke/mm	180
Displacement/ cm^3	905
Compression ratio	10
Combustion chamber	Disk-shape
Ignition method	Inductive spark
Spark electrode position	Center

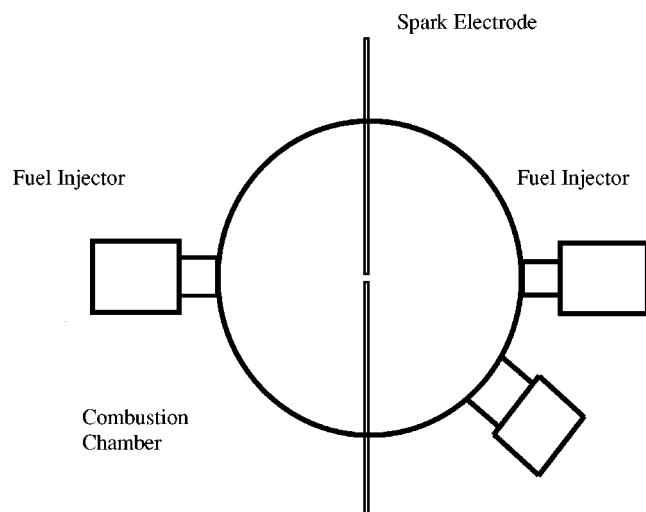


Fig. 1 Schematic of combustion chamber

Experimental Procedure. Natural gas was used in the experiment and its composition is listed in Table 2. Bomb dry air was used to remove the effect of humidity. The initial cylinder air pressure and temperature before compression were set at the atmospheric condition which varied during the days when the experiments were conducted. The daily variation was less than 2%. Figure 2 shows the history of cylinder pressure from the start of compression. The piston reaches the top dead center (TDC) at about 55 ms after the start of compression and is retained at TDC by a stopping mechanism. A fairly rapid pressure decline over the next 3 ms after 55 ms is due to the movement of the piston since it slightly overpasses TDC and returns back to TDC position to activate the stopping mechanism. The combustion then takes place under a constant volume condition. Combustion products were analyzed by using an FTIR analyzer (Horiba MEXA4000FT) which can measure 11 species of hydrocarbons with their carbon from C_1 to C_7 . Since it was determined that most of the unburned hydrocarbons in the experiment are unburned methane (>95%), it is reasonable to use unburned methane in the analysis of combustion efficiency due to unburned hydrocarbons.

Setting of Fuel-Injection Timing. The ignition timing was kept constant at 80 ms after the start of compression. In order to reveal the effect of fuel-injection timing relative to ignition timing on the combustion characteristics, three kinds of fuel-injection timing setting were selected. In the first setting, the fuel-injection start timing was kept constant at 60 ms from the start of compression as shown in Fig. 3(a). In this case, ignition timing always falls after the end of injection. In the second setting, the injection end timing was set at 80 ms after the start of compression (the same timing as the ignition timing) as shown in Fig. 3(b). In this case, ignition timing corresponds to the injection end timing. In the third setting, the fuel injection start timing was held constant at 75 ms, where the ignition timing always falls in the duration of the fuel injection as shown in Fig. 3(c). By applying these three injection timing settings, the authors expected to see the effect of charge stratification and turbulence behavior produced by the fuel jet.

Table 2 Composition of CNG used in this study

Manufacturer	Nippon Natural Gas Co., Ltd.
CH ₄	99.128
CO ₂	0.761
N ₂	0.094
C ₂ H ₆	0.011
O ₂	0.006

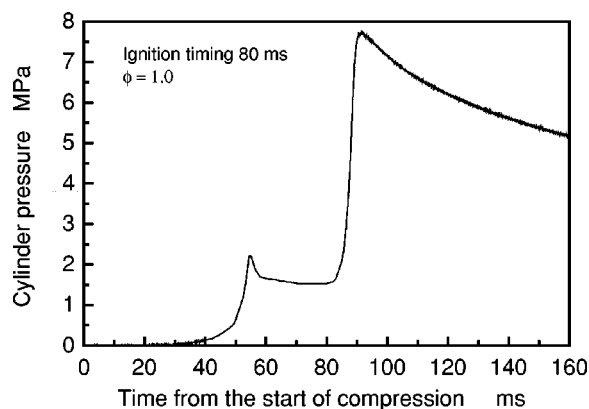


Fig. 2 Pressure history from start of compression

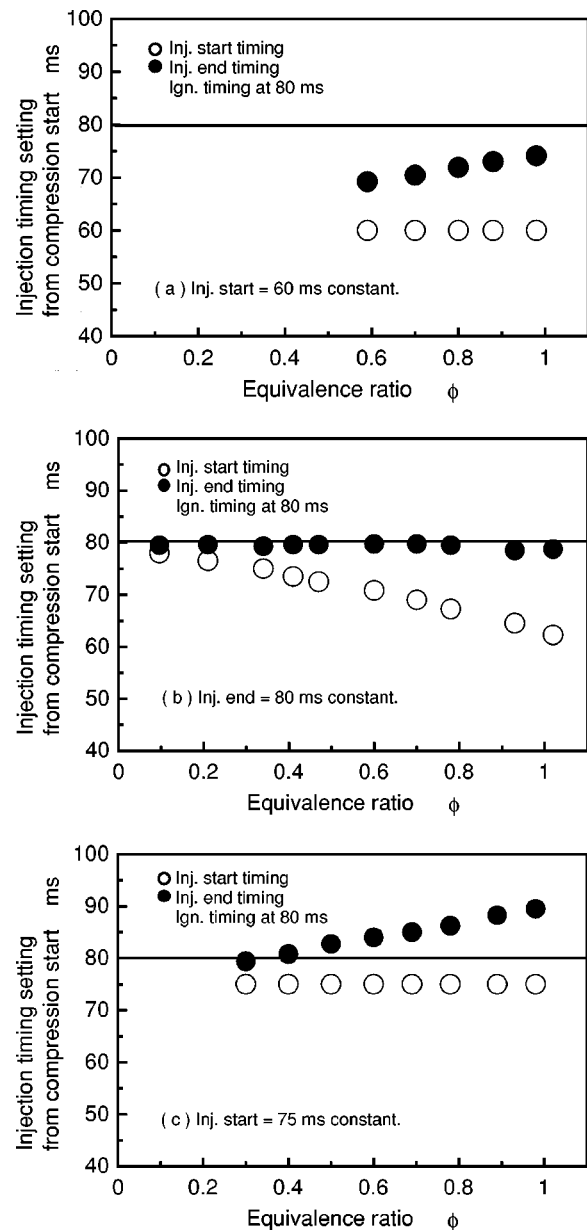


Fig. 3 Injection timing setting from start of compression

Results and Discussions

Definition of Combustion Parameters. Combustion duration is commonly used to describe combustion characteristics, where the initial, the main, the late, and the total duration of combustion are widely used. The initial duration is defined as the time interval from spark discharge to 10% of pressure rise due to combustion. Main duration is the time interval from 10% to 90% of pressure rise due to combustion. Late duration is the time interval from 90% to 100% of pressure rise due to combustion. The total duration is the sum of the three durations defined above.

Heat Release Analysis Using a Single Zone Model. Since the detail of a single zone model is also available in the previous study, [20], a brief description is given here. The term of the rate of heat transfer dQ_w/dt (kJ/s) is generally expressed as

$$\frac{dQ_w}{dt} = h_c \cdot A [(T - T_w) + 8.5 \times 10^{-11} (T^4 - T_w^4)], \quad (1)$$

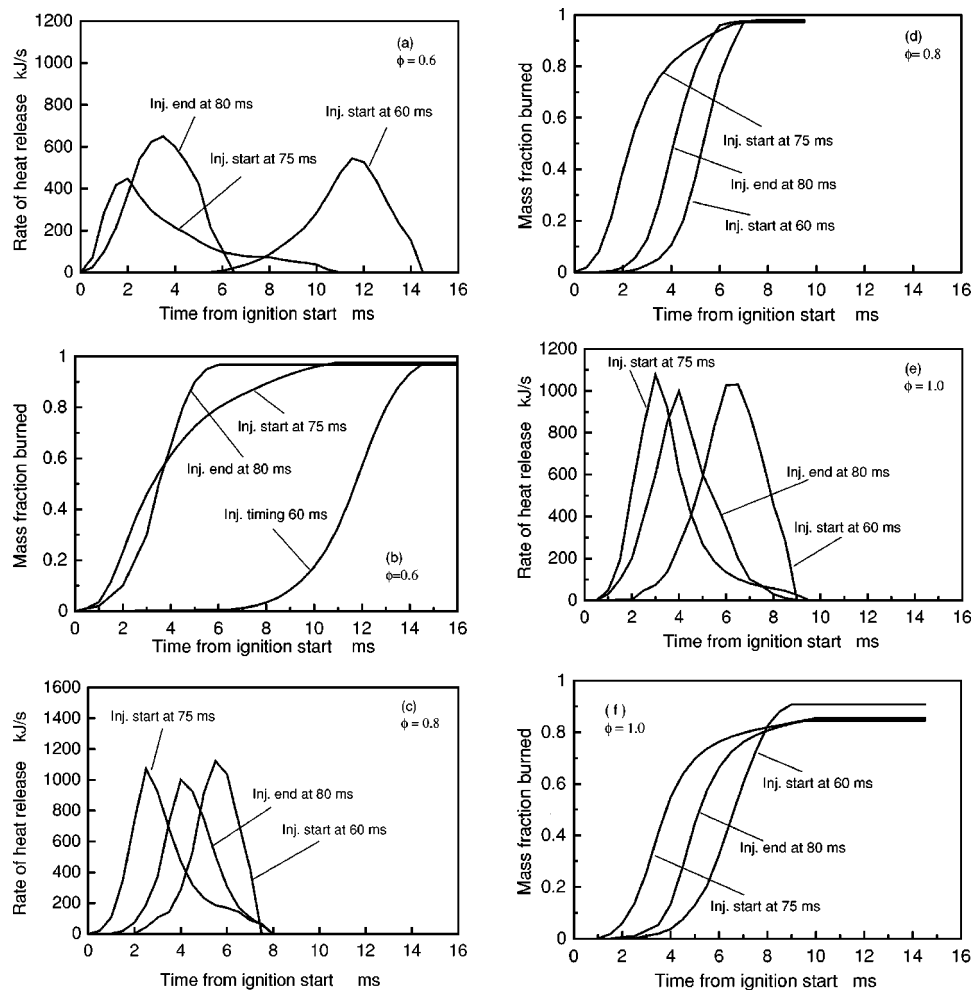


Fig. 4 Effect of injection timing relative to ignition timing on heat release rate and mass fraction burned

where A (m^2) is surface area, T_w (K) is wall temperature, and T (K) is gas temperature.

The heat transfer coefficient h_c ($\text{kW}/\text{m}^2\cdot\text{K}$) was given by

$$h_c = 0.49 \cdot k \cdot R_e^{0.7} / D \quad (2)$$

where $k = \mu \cdot C_p / 0.7$, $R_e = \rho \cdot D \cdot C_m / \mu$, $\mu = 3.3 \times 10^{-7} T^{0.7}$, and $C_m = 3.6$ m/s (C_m was estimated by considering RCM conditions which correspond to the engine conditions at 600 rpm), and the constant 0.49 was determined on the basis of the pressure decrease history after combustion.

Effect of Fuel Injection Timing on Heat Release Pattern.

The rate of heat release and the mass fraction burned for three types of injection timing settings are given in Fig. 4 at the equivalence ratio of 0.6, 0.8, and 1.0. The methods and equations for calculating the rate of heat release and mass fraction burned have been described in detail in the authors' previous work, [20].

When the injection was relatively early (injection start timing at 60 ms), the heat release pattern showed a slower burn in the initial stage and a faster burn in the late stage. There would be two factors causing this behavior: one is the degree of charge stratification of the mixture; the other is the decaying of the turbulence generated by the fuel jet. Since the injection timing was earlier, decaying of the formed turbulence and weaker charge stratification were considered to occur in the cylinder, resulting in a heat release pattern similar to that of flame propagation in the homogeneous mixture.

In contrast to this, when the injection timing was relatively late (injection start timing at 75 ms), the heat release rate showed a faster burn in the initial stage and a slower burn in the late stage, which is contrary to the early injection case. This pattern reminds us of diesel combustion in which two kinds of heat release process are usually identified: pre-mixed phase at the initial, and diffusive phase at the subsequent. Thus, for the case of late injection a mixture formed in the initial stage can burn rapidly due to strong turbulence and then in the subsequent stage the fuel probably burns like a diffusion flame.

The shortest combustion duration was realized at the injection end timing of 80 ms (the same timing as the ignition timing), which means the intermediate case of injection timing between the early and late injection. In terms of combustion promotion, this would give the optimum mixture formation by taking advantage of each case, the faster burn at the initial stage due to moderately strong turbulence and the faster burn at the late stage due to the moderately proceeding mixture formation. This phenomenon tends to be more obvious when decreasing the equivalence ratio. This would be due to the fact that contribution of the charge stratification to this optimizing behavior becomes greater with decreasing the equivalence ratio. Thus the main feature of the effect of fuel injection timing can be explained by the fuel-air mixing (charge stratification) and turbulence produced. Furthermore, it is shown that fuel injection timing can be an important parameter to control the combustion characteristics in natural gas direct-injection combustion. It was also found that the mass fraction

burned at the equivalence ratio of 1.0 dropped. This is due to the obvious increase in CO formation, since the excessive stratification of mixture might have been present at this large equivalence ratio, which is also consistent with the results of combustion efficiency analysis shown in a later section.

Effect of Injection Timing Relative to Ignition Timing on Combustion Duration. The initial, the main, the late and the total combustion durations plotted against the equivalence ratio are given in Fig. 5. The injection timing has a remarkable effect on the lean combustion limit. For the first setting of injection timing (injection start at 60 ms), the lean combustion limit is only $\phi=0.6$, and it is 0.3 at the third setting of injection timing (injection start at 75 ms). The second setting of injection timing (injection end at 80 ms) gives the lean limit of 0.1, which is pretty low considering that it is usually around 0.3 for a production gasoline direct-injection engine.

As for the initial combustion duration shown in Fig. 5(a), the dependence on equivalence ratio can be classified into two types: one is the case of early injection at the injection start timing of 60 ms where the initial duration is sensitive to equivalence ratio and increases steeply with the decrease of equivalence ratio, the other is the case of intermediate and late injection timing where the initial duration decreases monotonically with the decrease of equivalence ratio. In the latter case, the absolute value for late injection timing case is slightly lower than the intermediate injection timing case. These findings support the statement that for early injection the combustion characteristics are similar to that of homogeneous charge combustion, which can also be recognized in the similarity of the initial combustion duration dependence on equivalence ratio as shown in Fig. 5(a). As for the case of less time interval between the injection and ignition, the initial flame can grow rapidly due to the combustion of a richer mixture formed with the higher level of turbulence generated by the injected fuel jet. This would also be the main reason why the duration of the initial flame development is reduced by decreasing the time interval between the injection and ignition for the intermediate and late injection timing cases.

As for the main combustion duration shown in Fig. 5(b), although there is no remarkable difference, the duration for the intermediate case of the injection end timing of 80 ms almost takes a minimum. This is also due to the optimum mixture formation by taking advantage of the merits each case, a faster burn at the initial stage due to moderately strong turbulence and a faster burn at the late stage due to the moderately proceeding mixture formation.

As for the late combustion duration shown in Fig. 5(c), in the range of equivalence ratio between 0.5 to 0.9 late injection has a longer duration due to excessive charge stratification, which was also shown by the heat release pattern in Figs. 4(a) and (c).

Effect of Fuel-Injection Timing Relative to Ignition Timing on Combustion Products. Combustion products for these three injection timing settings are shown in Fig. 6. Unburned methane is not affected by the injection-timing settings and has little variation with equivalence ratio at $\phi < 0.8$ and show a slight increase with equivalence ratio at $\phi > 0.8$.

CO remains almost zero for all injection timing settings when equivalence ratio is less than 0.8, and it increases steeply with the increase in equivalence ratio at greater than 0.8. In terms of the injection timing, late timing of 75 ms produces relatively higher CO concentration than the other two cases. As shown previously, this must also be due to the excessive stratification.

NO_x reaches its peak value at the range of the equivalence ratio from 0.6 to 0.8 regardless of the fuel injection timings. NO_x reduces steeply with the increase in equivalence ratio in the range of greater than 0.8. Formation of NO_x is generally determined by oxygen concentration, equivalence ratio, the temperature history, and the residence time at high temperature. Excessive stratification would have caused less oxygen concentration in the combustion

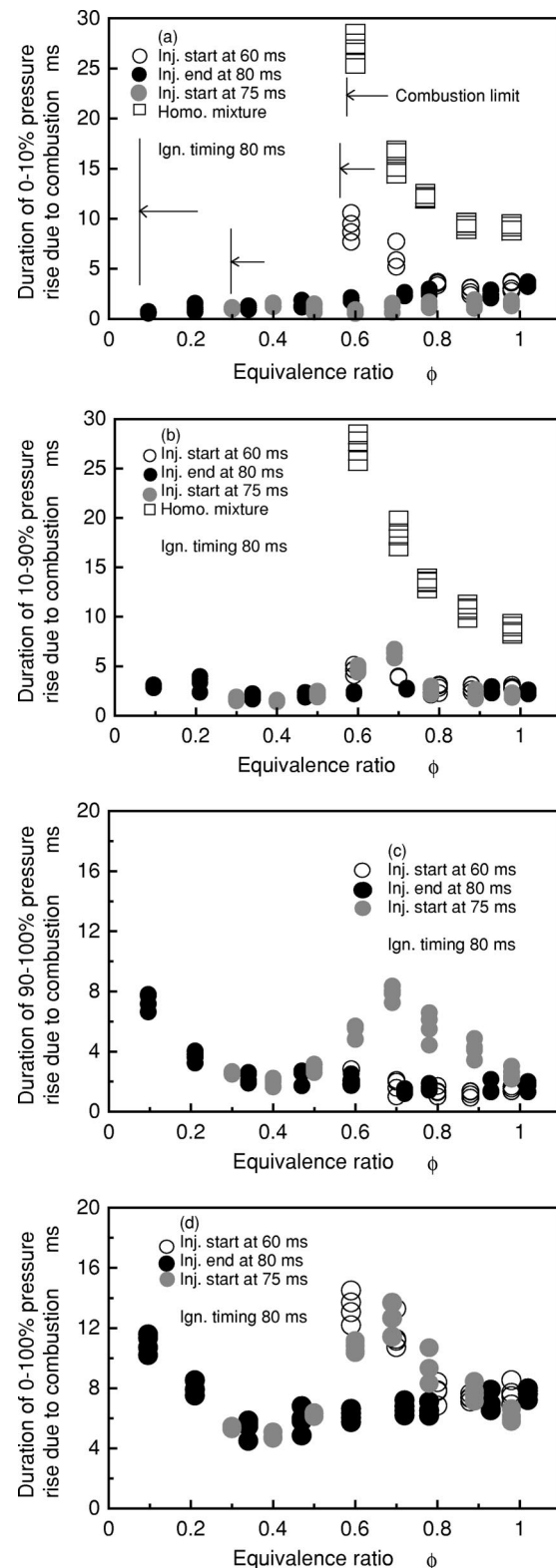


Fig. 5 Variation of combustion durations with injection timing relative to ignition timing

tion zone, which is related to the reduction of NO_x. It could also be determined that the mean temperature was not increased at the equivalence ratio greater than 0.8 where excessive stratification might have occurred. This was verified by the analysis of the maximum pressure rise due to combustion at the equivalence ratio

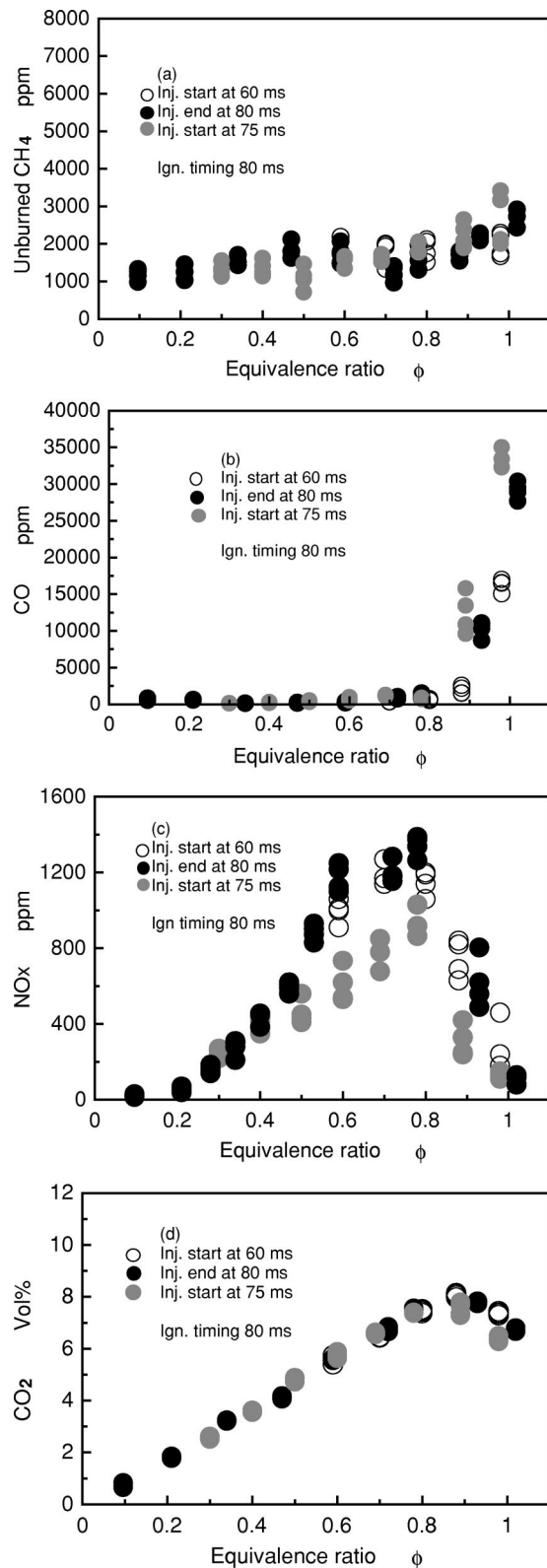


Fig. 6 Effect of injection timing relative to ignition timing on combustion products

greater than 0.8. Thus the restriction of temperature increase in the combustion at excessively stratified condition can be attributed to the decrease of NO_x .

This is a contrasting behavior to that of CO concentration and is caused by the charge stratification at larger equivalence ratio. Late

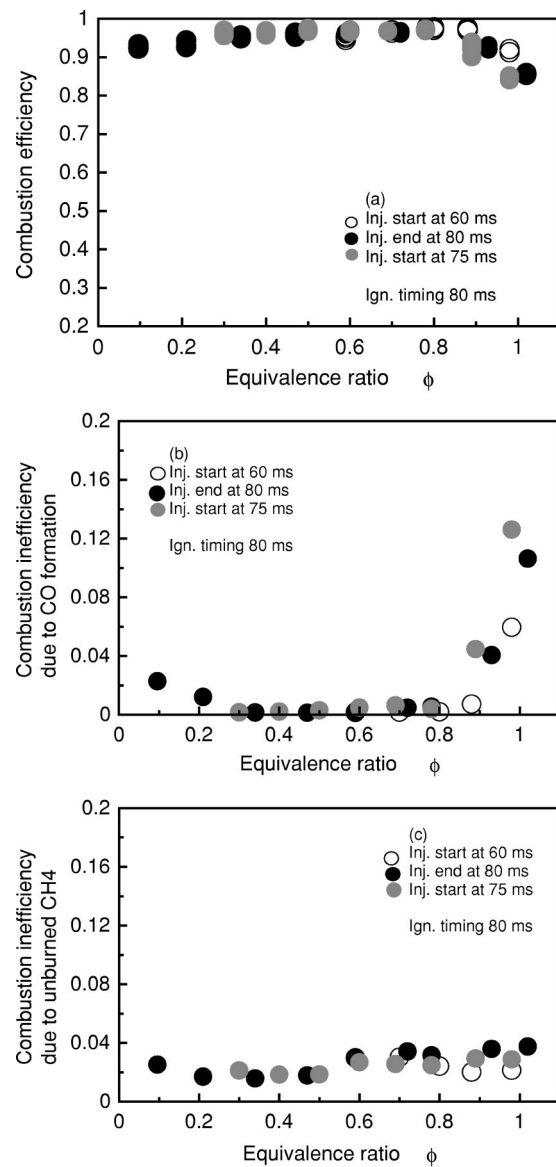


Fig. 7 Effect of injection timing relative to ignition timing on combustion efficiency

injection produces relatively lower NO_x at the equivalence ratio greater than 0.5, while CO level is higher for the latest injection at high equivalence ratio greater than 0.9. Thus the region where the NO_x and CO tradeoff exists is rather narrow, and between the equivalence ratio of 0.5 to 0.9 there is no tradeoff with low NO_x level suggesting other effect such as an internal EGR. This probable effect of internal EGR in the case of late injection would be worthwhile to verify, since there might be strong mixing between burned products and fuel jet due to the injection timing and the injector arrangement. There is a possibility to solve the NO_x problem accompanying any type of stratified combustion.

CO_2 concentration is not affected by the injection timing setting at equivalence ratio less than 0.8. In the range of equivalence ratio greater than 0.8, however, CO_2 shows a consistent behavior with that of CO. This is also reasonable since unburned methane was almost the same level regardless of the equivalence ratio. Thus, the difference in CO_2 for three cases of injection timing settings is caused by their difference in CO concentration when equivalence ratio is over 0.8.

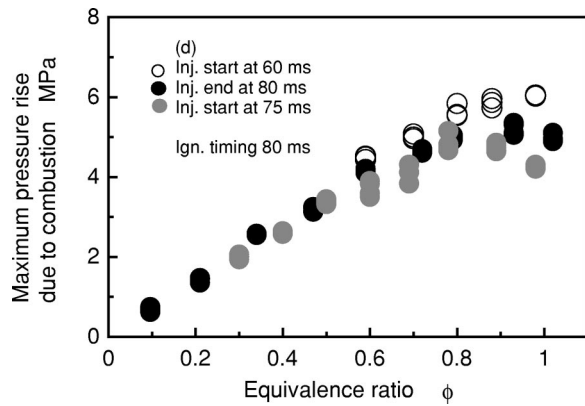


Fig. 8 Effect of injection timing relative to ignition timing on maximum pressure rise due to combustion

Effect of Fuel Injection Timing Relative to Ignition Timing on Combustion Efficiency. The estimation of combustion efficiency η_c from the exhaust gas components is given by Eq. (3) in Heywood's book, [14].

$$\eta_c = \frac{H_R(T_A) - H_P(T_A)}{m_f \cdot H_u}, \quad (3)$$

where $H_R(T_A) - H_P(T_A) = \sum_{i, \text{reactants}} n_i \Delta h_{f,i} - \sum_{i, \text{products}} n_i \Delta h_{f,i}$; n_i ; the number of moles of species i in the reactants or products, $\Delta h_{f,i}$; the standard enthalpy of formation of species i at ambient temperature T_A , $m_f \cdot H_u$; the amount of total fuel energy released by combustion.

The effect of unburned methane and CO on the combustion inefficiency ζ_{CH_4} and ζ_{CO} can also be derived from the above formula as follows:

$$\zeta_{\text{CH}_4} = \frac{0.5\phi + 4.773}{0.5\phi} \cdot \text{CH}_4(\text{ppm}) \times 10^{-6}, \quad (4)$$

$$\zeta_{\text{CO}} = \frac{\text{CO}(\text{ppm}) \cdot (0.5\phi + 4.773) \cdot (\Delta h_{f,\text{CO}}^0 - \Delta h_{f,\text{CO}_2}^0) \times 10^{-6}}{0.5\phi \cdot H_u}, \quad (5)$$

where $\Delta h_{f,\text{CO}_2}^0$ and $\Delta h_{f,\text{CO}}^0$ are the standard enthalpy of formation of CO_2 and CO at 25 °C, respectively.

As shown in Fig. 7(a), the calculation showed that combustion efficiency is maintained at a value greater than 0.92 in the range of equivalence ratio from 0.1 to 0.9, although it deteriorates at the equivalence ratio greater than 0.9. The formation of CO is the main contribution to the combustion inefficiency, which is due to charge stratification as shown in Fig. 7(b). Combustion inefficiency caused by unburned methane shows almost the same value regardless of the equivalence ratio. This means that 2 to 4% of the fuel always remains unburned and this fraction does not vary with the amount of fuel injected. This then suggests that a small percentage of fuel will not burn and it probably comes from the so-called bulk quenching at the periphery of fuel jet. Thus, it is shown that at lower equivalence ratio the combustion efficiency is satisfactorily high, and at higher equivalence ratio it deteriorates due to the CO formation which resulted in the decrease of the pressure rise due to combustion as shown in Fig. 8. Therefore, increasing the time interval between the injection and ignition can improve the combustion efficiency at higher equivalence ratio.

Conclusions

Effect of fuel injection timing relative to ignition timing on natural-gas direct-injection combustion was studied by using a rapid compression machine. The results obtained are summarized as follows.

1. The heat release pattern of early injection showed a slower burn in the initial stage and a faster burn in the late stage, which is similar to that of flame propagation of a premixed gas. In contrast to this, the heat release pattern of late injection showed a faster burn in the initial stage and a slower burn in the late stage, which is similar to that of diesel combustion including both pre-mixed and diffusion combustion phase. No clear evidence is given for both combustion patterns at this stage.

2. The shortest combustion duration realized at the setting of the injection end timing was the same as the ignition timing over the wide range of equivalence ratio. Early injection leads to a longer duration of the initial combustion, whereas late injection leads to a longer duration of the late combustion. The degree of charge stratification and the intensity of turbulence generated by fuel jet are considered to be responsible for this behavior as well as that of the heat release pattern stated in 1. Thus, there is an optimum injection timing at which the combustion duration is minimized and the heat release is fastest, which must be characteristic for CNG spark-ignition stratified-combustion.

3. Late injection produces relatively lower NO_x at the equivalence ratio greater than 0.5, while CO level is higher for the latest injection at the high equivalence ratio greater than 0.9. Thus the region where the NO_x and CO tradeoff exists is rather narrow, and between the equivalence ratio of 0.5 to 0.9 there is no tradeoff with low NO_x level suggesting other possible effects such as internal EGR. Combustion efficiency was maintained at a high value over a wide range of equivalence ratio. Considering that no smoke was detected over the whole range of experimental conditions, CNG stratified combustion has a great potential for practical use.

Acknowledgments

This study was supported by Nissan Motor Co. Ltd., and carried out under the cooperative research program. The authors also acknowledge Mr. Hiroyuki Watanabe, a student of Gunma University, for his help with the experiment and preparation of the manuscript.

References

- [1] Ben, L., Charnay, G., and Truquet, R., 1999, "Influence of Air/Fuel Ratio on Cyclic Variation and Exhaust Emission in Natural Gas SI Engine," SAE Paper No. 992901.
- [2] Hassaneen, A. E., Varde, K. S., and Bawady, A., 1998, "A Study of the Flame Development and Rapid Burn Durations in a Lean-Burn Fuel Injected Natural Gas S.I. Engine," SAE Paper No. 981384.
- [3] Liu, Z., and Karim, G. A., 1997, "Simulation of Combustion Process in Gas Fueled Diesel Engines," Proc. Inst. Mech. Eng., **211**(A2), pp. 159–169.
- [4] Roethlisberger, R. P., Leyland, G., and Favrat, D., 1998, "Study of a Small Size Cogeneration Gas Engine in Stoichiometric and Lean Burn Modes: Experimentation and Simulation," SAE Paper No. 982451.
- [5] Edward, R., Rosca, R., and Gaiginschi, R., 1999, "Influence of Fuel Injection Timing Over the Performances of a Direct Injection Spark Ignition Engine," SAE Paper No. 990174.
- [6] Brehob, D. D., Stein, R. A., and Haghgoie, M., 1998, "Stratified-Charge Engine Fuel Economy and Emission Characteristics," SAE Paper No. 982704.
- [7] Stovell, C., Matthews, R. D., and Henry K. N., 1998, "Analysis of Mixture Formation of Direct Injection Gasoline Engine," SAE Paper No. 980157.
- [8] Kano, M., Saito, K., and Basaki, M., 1998, "Emissions and Fuel Economy of a 1998 Toyota With a Direct Injection Spark Ignition Engine," SAE Paper No. 981462.
- [9] Green, R. K., and Zavier, C. C., 1992, "Charge Stratification in a Spark Ignition Engine," Proc. Inst. Mech. Eng., **206**(A1), pp. 59–64.
- [10] Rousseau, S., Lemoult, B., and Tazerout, M., 1999, "Combustion Characteristics of Natural Gas in a Lean Burn Spark-Ignition Engine," Proc. Inst. Mech. Eng., **213**(D5), pp. 481–489.
- [11] Jones, A. L., and Evans, R. L., 1985, "Comparison of Burning Rates in a Natural Gas Fueled Spark Ignition Engine," ASME J. Eng. Gas Turbines Power, **107**, pp. 908–913.
- [12] Pirouzpanah, V., and Kashani, B. O., 1999, "Prediction of Major Pollutants Emission in Direct Injection Dual-Fuel Diesel and Natural-Gas Engines," SAE Paper No. 990841.
- [13] Naber, J. D., 1994, "Effect of Natural Gas Composition on Ignition Delay Under Diesel Conditions," Combust. Flame, **99**, pp. 192–200.
- [14] Heywood, J. B., 1988, *Internal Combustion Engine Fundamentals*, McGraw-Hill, New York.

- [15] Obokata, T., Ishima, T., Shiga, S., Egurao, Y., Matsuda, T., Murakami, A., Long, W., Yang, X., and Liu, C., 2001, "Characteristics of Premixed Combustion Type Diesel Engine Using Hollow Cone Spray," *Proc. of ASME 2001 Fall Technical Conference*, **2**, ASME, New York, pp. 119–126.
- [16] Shiga, S., Kobayashi, K., Hayakawa, E., Matsuura, T., Nakamura, H., Ishima, T., and Obokata, T., 2001, "Effect of Injection Rate on the Performance and Emissions of a Direct-Injection Gasoline Engine in Comparison With the Port-Injection Operation," *Proc. of ASME 2001 Fall Technical Conference*, **1**, ASME, New York, pp. 115–120.
- [17] Agarwal, A., and Assanis, D. N., 2000, "Multi-Dimensional Modeling of Natural Gas Ignition, Combustion and Pollutant Formation in Direct Injection Engines," SAE Paper No. 201839.
- [18] Goto, Y., 1999, "Mixture Formation and Ignition in a Direct Injection Natural Gas Engine," *JSME Int. J., Ser. B*, **42**(2), pp. 268–274.
- [19] Brett, L., Machamara, J., Musch, P., and Simmie, M., 2001, "Simulation of Methane Autoignition in a Rapid Compression Machine With Creviced Pistons," *Combust. Flame*, **124**, pp. 326–329.
- [20] Shiga, S., Ueda, T., and Huang, Z., 2002, "A Study on the Combustion and Emission Characteristics of Compressed Natural Gas Direct Injection Stratified Combustion by Using a Rapid Compression Machine," *Combust. Flame*, **129**, pp. 1–10.

Statistical Correlation Between the Crankshaft's Speed Variation and Engine Performance—Part I: Theoretical Model

D. Taraza

Mechanical Engineering Department,
Wayne State University,
5050 Anthony Wayne Drive,
Detroit, MI 48202

The goal of this two-part paper is to develop a methodology using the variation of the measured crankshaft speed to calculate the mean indicated pressure (MIP) of a multicylinder engine and to detect cylinders that are lower contributors to the total engine output. Both the gas pressure torque and the crankshaft's speed are, under steady-state operating conditions, periodic functions of the crank angle and may be expressed by Fourier series. For the lower harmonic orders, the dynamic response of the crankshaft approaches the response of a rigid body and that makes it is possible to establish correlations between the amplitudes and phases of the corresponding harmonic orders of the crankshaft's speed and of the gas pressure torque. The inherent cycle-to-cycle variation in the operation of the cylinders requires a statistical approach to the problem. The first part of the paper introduces the statistical model for a harmonic component of the gas pressure torque and determines the correlation between the amplitudes and phases of the harmonic components of the gas pressure torque and the MIP of the engine. In the second part of the paper the statistical model is used to calculate the MIP and to detect deficient cylinders in the operation of a six-cylinder four-stroke diesel engine. [DOI: 10.1115/1.1563244]

Introduction

The cycle-to-cycle variations in the contribution of each cylinder to the total engine output are small for a healthy operating engine but, if something starts to deteriorate in the fuel injection system, cylinder sealing or ignition system (for spark ignition engines) the variation in the contribution of the different cylinders could grow larger with a negative impact on fuel consumption and exhaust emissions. In order to monitor the proper operation of the engine it is important to have a reliable diagnosis tool that could determine in real time the average engine torque and detect if all cylinders are contributing fairly uniform to the resultant engine torque.

The development of miniature torque sensors for the direct measurement of the output torque implies complex technical solutions, [1,2]. Also, they require additional space, which in many applications is not available and their reliability is questionable. In contrast, the crankshaft's speed may be easily and accurately measured and it is already a basic parameter in the engine controls. The use of the crankshaft speed for diagnosis purposes is very attractive and extensive research was carried out in this direction, especially in connection with the detection of misfiring cylinders in spark ignition engines.

The approaches used in detecting misfiring cylinders range from pattern recognition methods, [3,4], to elaborate methods based on the dynamic model of the crankshaft and the harmonic structure of the engine torque, [5–9]. Pure stochastic methods, [10,11], have been also developed to estimate the torque of individual cylinders by determining cross-correlation functions between the cylinder pressure and a number of parameters of the crankshaft's speed. In a review paper, [12], Williams summarizes the pulses and minuses of such diagnostic methods.

Because the crankshaft is an elastic structure, it is subjected to torsional vibrations that significantly distort its dynamic response

making it difficult to correlate the measured crankshaft speed with the gas pressure torque. The complex torsional motion of the crankshaft may be described by modal analysis as the superposition of the responses of its natural vibration modes to each harmonic component of the torque, [13]. Because the higher modes have very high natural frequencies, only the rigid-body mode and the first and second natural modes play a significant role in determining the torsional motion of the crankshaft, [13]. The method developed in this paper takes advantage of the fact that, at the low frequency of the lower harmonic components of the gas pressure torque, the contribution of the first and second natural modes becomes negligible and the dynamic response of the crankshaft approaches the response of a rigid body. Measuring the crankshaft speed at the flywheel, which is usually situated close to the node of the first natural mode, further reduces the influence of the higher modes. In this case it is possible to establish a direct correlation between the same harmonic order components of the crankshaft's speed and of the gas pressure torque.

The establishment of such a correlation is based on a vector interpretation of the harmonic components of the cylinder torque and the relative positions of these vectors for the different harmonic orders in a multicylinder engine. Because the cylinder torque is subjected to a random variation, a statistical model of the vector representing a harmonic component of the gas pressure torque is developed. The model is used to determine the characteristics of the vector representing a harmonic component of the resultant engine torque corresponding to different malfunctions in the operation of the cylinders. Replacing the resultant gas pressure torque by the MIP and the harmonic components of the cylinder torque by the corresponding harmonics of the tangential gas pressure (TGP) increase the generality of the problem. This approach allows the use of generic data found in the literature, when measured values are not available, to obtain correlations between the amplitudes and phases of the harmonic components of the TGP and the MIP of the engine.

In conjunction with the rigid-body model of the crankshaft these correlations will be used in the second part of the paper to estimate the MIP and detect deficient cylinders.

Contributed by the Internal Combustion Engine Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the ICE Division, September 2001; final revision received by the ASME Headquarters, July 2002. Associate Editor: D. Assanis.

The Harmonic Structure of the Engine Torque

The torque produced by a cylinder has two components, the gas pressure torque (GPT) and the reciprocating inertia torque (RIT). The GPT depends, almost exclusively, on the engine load and could vary from cylinder to cylinder even under steady-state operating conditions. The RIT depends only on engine speed and is fairly uniform for all cylinders of a multicylinder engine. Under steady-state operating conditions, the total torque (T_i) corresponding to a given cylinder i , may be considered a periodic function of time (crank angle) and expressed as a Fourier series. The following relation expresses the total cylinder torque for a four-stroke engine:

$$T_i = T_{0_i} + \sum_{j=1}^K \left[A_{j/2} \cos \frac{j}{2} \theta + B_{j/2} \sin \frac{j}{2} \theta \right]. \quad (1)$$

In this expression, T_{0_i} represents the mean value of the GPT, $A_{j/2}$ and $B_{j/2}$ are the cosine and sine terms of the order $k = j/2$ of the harmonic components of the total cylinder torque, θ is the crank angle, and K the highest number of harmonic components, necessary to represent the gas pressure torque (usually $K = 24$ for four-stroke engines). While the harmonic components of the GPT have both sine and cosine terms, the harmonic components of the RIT have only integer-order sine terms, [14]. Considering a lumped-mass model of the crankshaft and linear equations for the angular motion, the crankshaft's speed variation during an engine cycle may be quite accurately predicted by superposing the responses of its natural modes to each harmonic component of the total engine torque, [13,15]. The frequencies of the lower harmonic orders are usually far enough from the natural frequencies of the shafting and, for these frequencies, the influence of the higher modes becomes negligible. For a low-frequency excitation, the motion of the shafting is determined mainly by the rigid-body mode. In this case, under the action of a harmonically varying torque T_k ,

$$T_k = |T_k| \sin k\omega t, \quad (2)$$

the corresponding angular speed variation will be

$$\dot{\theta}_k = \frac{-|T_k|}{k\omega J_{\text{tot}}} \cos k\omega t = \frac{|T_k|}{k\omega J_{\text{tot}}} \sin \left(k\omega t - \frac{\pi}{2} \right) \quad (3)$$

where J_{tot} is the mass moment of inertia of the rotating shafts plus the corresponding mass moment of inertia of the reciprocating masses.

Equation (3) shows that the k th-order harmonic component of the crankshaft's speed lags the corresponding harmonic term of the engine torque by 90° . This equation may be used to correlate the amplitudes and phases of the lower harmonic components of the resultant torque and of the measured crankshaft speed.

Statistical Model of the Harmonic Component of the Gas-Pressure Torque (GPT)

A harmonic component of the GPT may be interpreted as the projection of a rotating vector on the crankshaft axis, [16]. For an arbitrary harmonic order k , the vector rotates k times faster than the crankshaft. The vectors corresponding to different cylinders will be phased according to the firing order of the engine. For example, Fig. 1 shows the phase angle diagrams for a four-stroke six-cylinder engine.

If all cylinders are equal contributors to the engine torque, the vectors in the phase angle diagrams have the same magnitude and are uniformly distributed at equal angular intervals. In reality, due to cycle-to-cycle variations or nonuniform cylinder operation, both magnitude and angular position may differ from the ideal situation shown in Fig. 1.

Considering an arbitrary cylinder i , the tip of the vector T_{k_i} interpreting the harmonic order k of the GPT would lay inside an ellipse of dispersion, having as half-axes $3\sigma_{T_{k_i}}$ and $3T_{k_i}\sigma_{\varphi_i}$ (Fig. 2).

The statistical model of the random vector T_{k_i} assumes a normal (Gaussian) distribution with standard deviations $\sigma_{T_{k_i}}$ for the vector magnitude and σ_{φ_i} for the angular position.

To simplify the calculations, a Cartesian system of coordinates is introduced. The two projections of the random vector T_{k_i} on the axes of the coordinate system are also normal distributed random vectors, having the following statistical parameters, [17–19]: mean values:

$$\begin{cases} \bar{T}_{k_{ix}} = \bar{T}_{k_i} \sin \bar{\varphi}_{k_i} \\ \bar{T}_{k_{iy}} = \bar{T}_{k_i} \cos \bar{\varphi}_{k_i} \end{cases} \quad (4)$$

and standard deviations:

$$\begin{cases} \sigma_{T_{k_{ix}}}^2 = \sigma_{T_{k_i}}^2 \sin^2 \bar{\varphi}_{k_i} + \bar{T}_{k_i}^2 \sigma_{\varphi_k}^2 \cos^2 \bar{\varphi}_{k_i} \\ \sigma_{T_{k_{iy}}}^2 = \sigma_{T_{k_i}}^2 \cos^2 \bar{\varphi}_{k_i} + \bar{T}_{k_i}^2 \sigma_{\varphi_k}^2 \sin^2 \bar{\varphi}_{k_i} \end{cases} \quad (5)$$

These random vectors, resulting from the projection of the same vector are correlated random variables having a coefficient of correlation:

$$\rho_{x_i y_i} = \frac{\tan 2\bar{\varphi}_{k_i}}{2} \frac{\sigma_{T_{k_{iy}}}^2 - \sigma_{T_{k_{ix}}}^2}{\sigma_{T_{k_{ix}}} \sigma_{T_{k_{iy}}}} \quad (6)$$

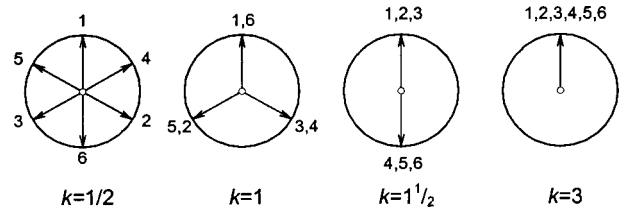


Fig. 1 Phase angle diagrams of the lower harmonic components of the gas pressure torque; four-stroke six-cylinder engine

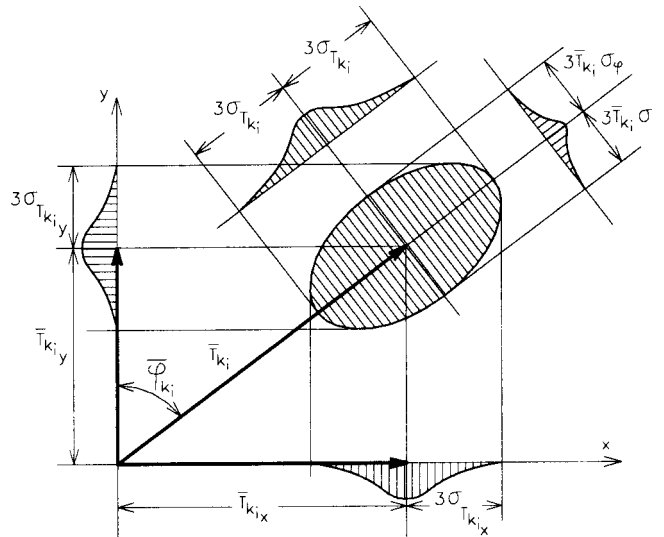


Fig. 2 Statistical distribution of the random vector representing the k th harmonic component of the GPT for cylinder i

and a covariance:

$$K_{x,y_i} = \rho_{x,y_i} \sigma_{T_{k_x}} \sigma_{T_{k_y}} = \frac{\sin 2\bar{\varphi}_{k_i}}{2} [\sigma_{T_{k_i}}^2 - \bar{T}_{k_i}^2 \sigma_{\varphi_k}^2]. \quad (7)$$

The resultant torque, representing the action of the k th harmonic component of the GPT for all N cylinders of the engine is found by determining its two components on the Ox and Oy -axes. These components result from the summation of independent random variables.

Several distinct situations may occur in the calculation of the resultant of the harmonic components of the cylinders' torque.

Symmetrical Phase-Angle Diagram

Uniform Contribution of All Cylinders. This is the typical situation for the half-order component $k=1/2$ (see Fig. 1). If the cylinders operate rather uniform, the mean values and the standard deviations for these harmonic components of the GPT will be the same for all cylinders:

$$\begin{aligned} \bar{T}_{k_i} &= \bar{T}_k & \sigma_{T_{k_i}} &= \sigma_{T_k} & \sigma_{\varphi_{k_i}} &= \sigma_{\varphi_k} \\ i &= 1, 2, \dots, N \end{aligned} \quad (8)$$

At the same time, due to the symmetry of the phase angle diagrams for $k=1/2$ and $k=1$:

$$\begin{aligned} \sum_{i=1}^N \sin \bar{\varphi}_{k_i} &= \sum_{i=1}^N \cos \bar{\varphi}_{k_i} = 0 \\ \sum_{i=1}^N \sin 2\bar{\varphi}_{k_i} &= \sum_{i=1}^N \cos 2\bar{\varphi}_{k_i} = 0. \end{aligned} \quad (9)$$

In this case the statistical parameters of the resultant torque vector are

- mean values:

$$\begin{aligned} \bar{T}_{k_x} &= \sum_{i=1}^N \bar{T}_{k_{i_x}} = \bar{T}_k \sum_{i=1}^N \sin \bar{\varphi}_{k_i} = 0 \\ \bar{T}_{k_y} &= \sum_{i=1}^N \bar{T}_{k_{i_y}} = \bar{T}_k \sum_{i=1}^N \cos \bar{\varphi}_{k_i} = 0 \end{aligned} \quad (10)$$

- standard deviations:

$$\begin{aligned} \sigma_{T_{k_x}}^2 &= \sum_{i=1}^N \sigma_{T_{k_{i_x}}}^2 = \sigma_{T_k}^2 \left(\frac{N}{2} - \frac{1}{2} \sum_{i=1}^N \cos 2\bar{\varphi}_{k_i} \right) \\ &+ \bar{T}_k^2 \sigma_{\varphi_k}^2 \left(\frac{N}{2} + \frac{1}{2} \sum_{i=1}^N \cos 2\bar{\varphi}_{k_i} \right) = \frac{N}{2} (\sigma_{T_k}^2 + \bar{T}_k^2 \sigma_{\varphi_k}^2) \\ \sigma_{T_{k_y}}^2 &= \sum_{i=1}^N \sigma_{T_{k_{i_y}}}^2 = \sigma_{T_k}^2 \left(\frac{N}{2} + \frac{1}{2} \sum_{i=1}^N \cos 2\bar{\varphi}_{k_i} \right) \\ &+ \bar{T}_k^2 \sigma_{\varphi_k}^2 \left(\frac{N}{2} - \frac{1}{2} \sum_{i=1}^N \cos 2\bar{\varphi}_{k_i} \right) = \sigma_{T_k}^2 = \sigma_{T_{k_x}}^2 \end{aligned} \quad (11)$$

and covariance:

$$K_{xy} = \sum_{i=1}^N K_{x,y_i} = \frac{1}{2} [\sigma_{T_{k_i}}^2 - \bar{T}_{k_i}^2 \sigma_{\varphi_k}^2] \sum_{i=1}^N \sin 2\bar{\varphi}_{k_i} = 0. \quad (12)$$

In this situation, the ellipse of dispersion reduces to a circle with the center in the origin of the coordinate system and the radius equal to three times the standard deviation σ_{T_k} . Thus, for uniform contribution of all cylinders, the resultant torque vector of the lowest harmonic order ($k=1/2$) will have a very small value and an arbitrary angular position. This situation is presented in Fig. 3.

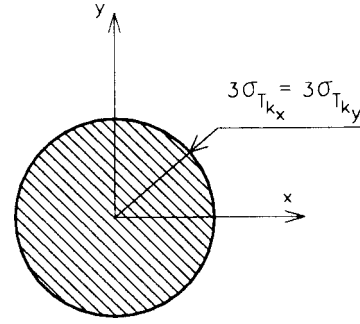


Fig. 3 The ellipse of dispersion for uniform contributions of all cylinders ($k=1/2$)

Nonuniform Contributions of the Cylinders. The lowest harmonic order $k=1/2$ is considered again and it is assumed that $N-1$ cylinders are contributing rather uniformly but, one cylinder contributes less than the others. In this case, the statistical parameters will be

$$\begin{aligned} \bar{T}_{k_i} &= \bar{T}_k & \sigma_{T_{k_i}} &= \sigma_{T_k} & \sigma_{\varphi_{k_i}} &= \sigma_{\varphi_k} \\ i &= 1, 2, \dots, N; & i &\neq S \end{aligned} \quad (13)$$

and, for the cylinder that contributes less,

$$\bar{T}_{k_s}, \quad \sigma_{T_{k_s}}, \quad \sigma_{\varphi_{k_s}}. \quad (14)$$

For the $N-1$ cylinders that contribute uniformly, the statistical characteristics may be obtained from Eqs. (10) to (12), which are taking the following form in this case:

$$(\bar{T}_{k_x})_{N-1} = \bar{T}_k \sum_{i=1}^{N-1} \sin \bar{\varphi}_{k_i} \quad (15)$$

$$(\bar{T}_{k_y})_{N-1} = \bar{T}_k \sum_{i=1}^{N-1} \cos \bar{\varphi}_{k_i}$$

$$(\sigma_{T_{k_y}}^2)_{N-1} = \sigma_{T_k}^2 \sum_{i=1}^{N-1} \cos^2 \bar{\varphi}_{k_i} + \bar{T}_k^2 \sigma_{\varphi_k}^2 \sum_{i=1}^{N-1} \sin^2 \bar{\varphi}_{k_i} \quad (16)$$

$$(K_{xy})_{N-1} = \frac{1}{2} (\sigma_{T_k}^2 - \bar{T}_k^2 \sigma_{\varphi_k}^2) \sum_{i=1}^{N-1} \sin 2\bar{\varphi}_{k_i}. \quad (17)$$

Due to the nonsymmetrical configuration of the $N-1$ vectors, all statistical parameters will be different from zero. The nonzero covariance indicates that the axes of coordinates are no more principal axes for the ellipse of dispersion. The angle γ between the principal axes of the ellipse of dispersion and the axes of coordinates is

$$\tan 2\gamma = \frac{-2(K_{xy})_{N-1}}{(\sigma_{T_{k_x}}^2)_{N-1} - (\sigma_{T_{k_y}}^2)_{N-1}} \quad (18)$$

and the values of the principal half-axes of the ellipse of dispersion are equal to three times the corresponding standard deviations:

$$\begin{aligned} \sigma_s^2 &= (\sigma_{T_{k_x}}^2)_{N-1} \cos^2 \gamma + (\sigma_{T_{k_y}}^2)_{N-1} \sin^2 \gamma + (K_{xy})_{N-1} \sin 2\gamma \\ \sigma_\eta^2 &= (\sigma_{T_{k_x}}^2)_{N-1} \sin^2 \gamma + (\sigma_{T_{k_y}}^2)_{N-1} \cos^2 \gamma - (K_{xy})_{N-1} \sin 2\gamma. \end{aligned} \quad (19)$$

The projections of the vector interpreting the deficient cylinder and their statistical parameters may be evaluated according to Eqs. (4) to (7), replacing the subscript i by s . Then, the statistical pa-

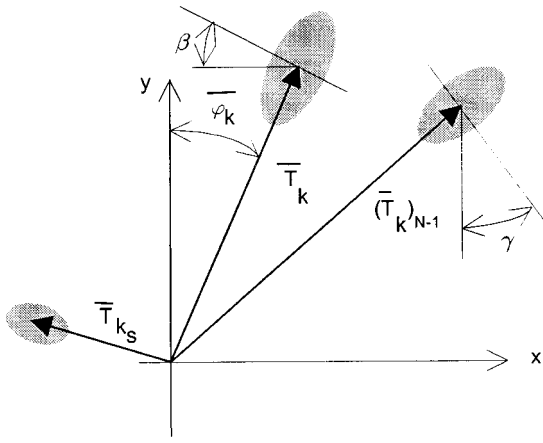


Fig. 4 Mean values and ellipses of dispersion for nonuniform cylinder operation

rameters of the resultant torque, corresponding to the harmonic order k and the contribution of all cylinders will be

- Mean values:

$$\begin{aligned}\bar{T}_{k_x} &= (\bar{T}_k)_{N-1} + \bar{T}_{k_{sx}} \\ \bar{T}_{k_y} &= (\bar{T}_k)_{N-1} + \bar{T}_{k_{sy}}\end{aligned}\quad (20)$$

- Standard deviations:

$$\begin{aligned}\sigma_{T_{k_x}}^2 &= (\sigma_{T_k}^2)_{N-1} + \sigma_{T_{k_{sx}}}^2 \\ \sigma_{T_{k_y}}^2 &= (\sigma_{T_k}^2)_{N-1} + \sigma_{T_{k_{sy}}}^2\end{aligned}\quad (21)$$

- Covariance:

$$K_{xy} = (K_{xy})_{N-1} + K_{x_{sy}} \quad (22)$$

Again, the covariance is different from zero and the principal axes of the ellipse of dispersion for the resultant torque of order k will be rotated with respect to the coordinate system by an angle β :

$$\tan 2\beta = \frac{-2K_{xy}}{\sigma_{T_{k_x}}^2 - \sigma_{T_{k_y}}^2} \quad (23)$$

The values of the principal half axes of the ellipse of dispersion for the k -order component of the total torque will be equal to three times the standard deviations:

$$\begin{aligned}\sigma_\xi^2 &= \sigma_{T_{k_x}}^2 \cos^2 \beta + \sigma_{T_{k_y}}^2 \sin^2 \beta + K_{xy} \sin 2\beta \\ \sigma_\mu^2 &= \sigma_{T_{k_x}}^2 \sin^2 \beta + \sigma_{T_{k_y}}^2 \cos^2 \beta - K_{xy} \sin 2\beta.\end{aligned}\quad (24)$$

In contrast with the case of uniform contribution of all cylinders, where the amplitude of the half-order torque vector is small and included in the ellipse of dispersion (see Fig. 3) in the case of a deficient cylinder the amplitude of the half-order torque vector grows larger and is subjected to a variation determined by the ellipse of dispersion (Fig. 4). Also, depending on the deficient cylinder, the resultant torque vector will lay in a definite phase domain.

Major Harmonic Orders

The major harmonic orders, for four-stroke engines, are multiples of the half of the number of cylinders. The first major harmonic order is $k=N/2$ (for a six-cylinder engine it will be $k=3$). The major harmonic components of the GPT have all the same phase (see Fig. 1) and their resultant is, theoretically, equal to N times the value corresponding to one cylinder. The RIT has

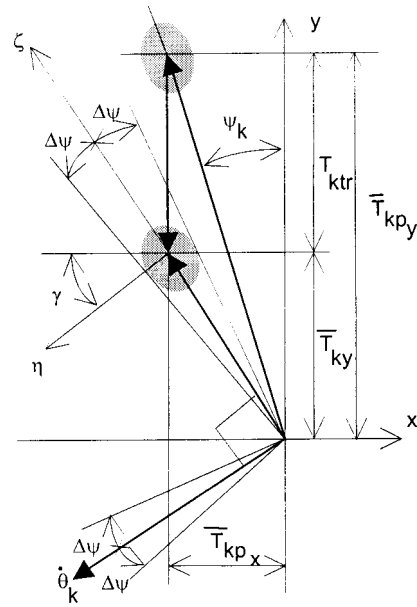


Fig. 5 The resultant of a major harmonic component of the engine torque

also a component on the lowest major harmonic order ($T_{k_{ir}}$), which must be added to the component of the GPT. The statistical scatter of the components of the RIT is much lower than the scatter of the GPT components and the RIT may be considered as constant in this analysis. If the statistical model is considered, the statistical parameters of the resultant torque are

$$\bar{T}_{k_x} = \bar{T}_{k_{px}} = N \times \bar{T}_k \sin \bar{\varphi}_k \quad (25)$$

$$\bar{T}_{k_y} = N \times \bar{T}_k \cos \bar{\varphi}_k + N \times T_{k_{tr}}$$

$$\sigma_{T_{k_x}}^2 = \frac{N}{2} [\sigma_{T_k}^2 (1 - \cos 2\bar{\varphi}_k) + \bar{T}_k^2 \sigma_{\varphi_k}^2 (1 + \cos 2\bar{\varphi}_k)]$$

$$\sigma_{T_{k_y}}^2 = \frac{N}{2} [\sigma_{T_k}^2 (1 + \cos 2\bar{\varphi}_k) + \bar{T}_k^2 \sigma_{\varphi_k}^2 (1 - \cos 2\bar{\varphi}_k)] \quad (26)$$

$$K_{k_{xy}} = \frac{N}{2} (\sigma_{T_k}^2 - \bar{T}_k^2 \sigma_{\varphi_k}^2) \sin 2\bar{\varphi}_k \quad (27)$$

The nonzero value of the covariance shows that the principal axes of the ellipse of dispersion are rotated with respect to the axes of coordinates by the angle γ :

$$\gamma = \frac{1}{2} \tan^{-1} \left(\frac{-2K_{k_{xy}}}{\sigma_{T_{k_x}}^2 - \sigma_{T_{k_y}}^2} \right) \quad (28)$$

This situation is shown in Fig. 5.

Correlation Between the Harmonic Components of the Gas Pressure Torque (GTP) and the Mean Indicated Pressure (MIP)

Due to the symmetry of its phase angle diagram (see Fig. 1), the lowest harmonic order ($k=1/2$) of the GPT is best suited to detect abnormal contributions of the cylinders to the total engine output. A significant reduction in the contribution of a cylinder will disturb the symmetry of the half-order vectors yielding a resultant vector, which will have a definite phase relationship with respect to the deficient cylinder.

The lowest major harmonic order ($k=N/2$) of the engine torque will have the tendency to level up the small phase differ-

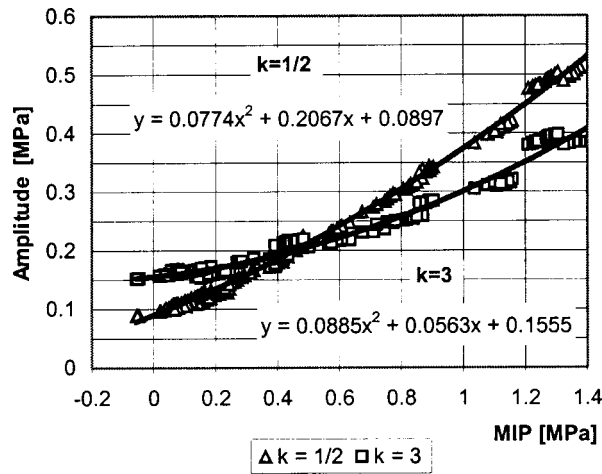


Fig. 6 The correlation between the amplitudes of the harmonic orders $k=1/2$ and $k=3$ of the tangential gas pressure and the MIP (measured values)

ences between the corresponding random vectors yielding a resultant that may be used to estimate the average value of the engine torque.

In order to correlate the corresponding harmonic components of the measured engine speed to the engine torque (engine load) a six-cylinder four-stroke direct injection diesel engine was investigated. A more general character of the measured data is obtained by replacing the harmonic components of the GPT by the corresponding tangential gas pressure (p_{T_k}) and the mean value of the gas pressure torque $\bar{T}_p$, by the MIP. These are specific parameters independent from the engine size:

$$p_{T_k} = T_k / \left(\frac{\pi D^2}{4} R \right) \quad (29)$$

and

$$MIP = 2 \pi \bar{p}_T = 8 \bar{T}_p / (D^2 R). \quad (30)$$

In these relations D represents the cylinder bore, R the crank radius, and $\bar{p}_T$ the mean value of the tangential gas pressure.

In order to establish correlations between these specific parameters the six-cylinder diesel engine equipped with pressure transducers on each cylinder was operated at different loads in a speed range 1200 to 1800 rpm. Figure 6 shows the results of the measurements and the corresponding correlations between the amplitudes of the harmonic components $k=1/2$ and $k=3$ of the TGP and the MIP of the corresponding cylinder.

The degree of generality of these correlations may be judged by comparison with generic data found in the literature, [15]. The marks in Fig. 7 represent the generic data for the third-order component of the TGP recommended for a small direct injection diesel engine, [15]. On the same figure the line represents the least-square fitting of the experimental data shown in Fig. 6. Inspection of Fig. 7 shows a good agreement between the measured data and the generic data given in the literature.

In a similar way correlations may be established between the phases of the harmonic orders of the TGP and the MIP. Figure 8 shows the correlation between the phases of the half ($k=1/2$) and third ($k=3$) harmonic orders of the TGP and the MIP.

The data in Figs. 6 and 8 are used to determine the mean values and standard deviations of the amplitudes and phases of the half harmonic orders of the TGP corresponding to mean values and standard deviations of the MIP. Once these parameters are determined, the same parameters of the half-order component of the GPT are calculated using relation (29) and the statistical model of the harmonic component of the GPT may be used to calculate the mean value and phase domain of the corresponding resultant

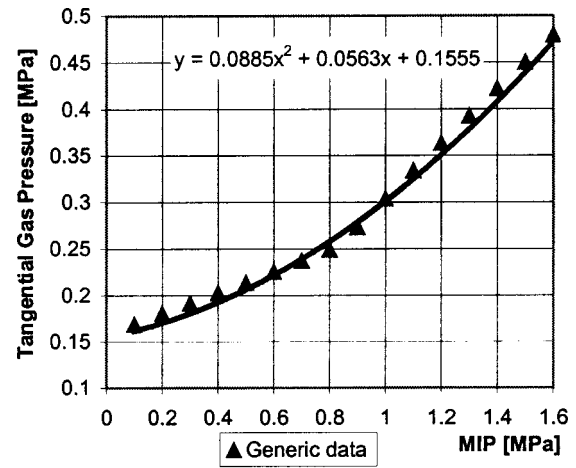


Fig. 7 Comparison between measured and generic data for the correlation TGP-MIP

torque vector. A deficient cylinder may be simulated and the expected phase domain of the half-order speed vector may be determined on the basis of the rigid-body model of the shafting. In reverse, from the phase domain of the measured half order speed vector, the deficient cylinder may be identified.

The major harmonic order may be treated in the same way. In this case the corresponding harmonic component of the RIT must be also calculated based on the average engine speed. For the six-cylinder engine, the third harmonic order of the RIT is ([14]):

$$T_{3_i} = 6 \left[m_{tr} (R \omega)^2 \left(-\frac{3}{4} \Lambda - \frac{9}{32} \Lambda^3 - \dots \right) \sin 3\theta \right] \quad (31)$$

where $\Lambda = R/L$, and L is the length of the connecting rod.

Using the statistical model, the mean value of the resultant torque and the dimensions of the ellipse of dispersion are determined. Then, based on relation (3) the mean value and standard deviations of the third-order component of the engine speed may be calculated. In reverse, the MIP of the engine may be estimated from the measured engine speed and the amplitude and phase of its third-order harmonic component.

The actual procedures of using the statistical model of the harmonic component of the GPT to calculate the MIP and determine

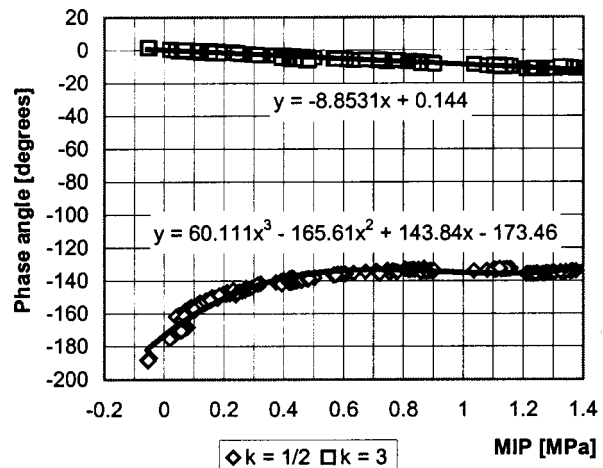


Fig. 8 The correlation between the phases of the harmonic orders $k=1/2$ and $k=3$ of the tangential gas pressure and the MIP (measured values)

if all cylinders are uniformly contributing to the total engine output are presented in detail in Part II of the paper.

Summary and Conclusions

It is very important to monitor, in real time, the healthy operation of a multicylinder engine and detect, in an early stage, the development of a malfunction. The diagnosis method must be simple and reliable and the measurement of the angular speed of the crankshaft may be easily and accurately performed. The main problem of correlating the crankshaft speed variations with the engine torque is created by the elasticity of the crankshaft and the torsional vibrations excited by the torque. In order to avoid the disturbances produced by torsional vibrations, the problem is restricted to the analysis of the correlations that could be established between the lower harmonic components of the engine torque and of the crankshaft speed.

At the frequencies of the lower harmonic components of each cylinder torque, the dynamic response of the crankshaft approaches the response of a rigid body and, for the same harmonic order the speed component will lag the torque component by almost 90°. For a multicylinder engine the vectors interpreting the harmonic components of the cylinder torques are either symmetrically distributed at equal angular intervals, or superimposed having all the same direction. Due to inherent cycle-to-cycle variations in the operation of the cylinders the torque of each cylinder fluctuates and, even for the symmetrical pattern of the torque vectors their resultant is different from zero. A statistical model of the vector interpreting the GPT of a cylinder is developed and used to calculate the mean value and phase domain of the resultant GPT vector.

For the lowest harmonic component of the GPT ($k=1/2$) the torque vectors are symmetrically disposed and their resultant should have a very low mean value and an arbitrary phase if all cylinders are equal contributors to the total engine output. If the symmetry is disturbed by a lower contribution of a cylinder, the resultant grows larger and so does also the amplitude of the half order component of the measured crankshaft speed. Based on the statistical model of the torque vector, the phase domain of the resultant GPT vector may be determined and correlated with the deficiency of each cylinder. In turn, from the phase domain of the measured half-order crankshaft speed component and the rigid-body model of the crankshaft the deficient cylinder can be identified.

For the major harmonic orders, for which all the GPT vectors have the same direction, the mean value of the resultant torque has a large value. This value can be correlated with the amplitude of the measured crankshaft speed on the basis of the rigid-body model of the crankshaft.

In conclusion, the half harmonic order of the measured crankshaft speed may be used to detect deficient cylinders and the first

major harmonic order to determine the average GPT or the MIP of the engine. The application of the model and method presented in the paper required the development of empirical correlations between the amplitudes and phases of the half order and of the first major orders for the experimental six-cylinder engine. A good agreement has been found between the correlations obtained from measurements on the experimental engine and the generic data given in the literature.

References

- [1] Klauber, R. D., Vigmstad, E. B., Van Gerpen, D., Sprague, F. P., and Reiter, F. J., 1992, "Miniature Magnetostrictive Misfiring Sensor," SAE Paper No. 920236.
- [2] Garshelis, I. J., Aleksonis, J. A., Jones, C. A., and Rotay, R. M., 1997, "Development of a Magnetoelastic Torque Transducer for Automotive Transmission Applications," SAE Paper No. 970605.
- [3] Rizzoni, G., 1989, "Diagnosis of Individual Cylinder Misfires by Signature Analysis of Crankshaft Speed Fluctuations," SAE Paper No. 89884.
- [4] Brown, T. S., and Neil, W. S., 1992, "Determination of Engine Cylinder Pressures From Crankshaft Speed Fluctuations," SAE Paper No. 920463.
- [5] Citron S. J., O'Higgins, J. E., and Chen, L. Y., 1989, "Cylinder by Cylinder Engine Pressure and Pressure Torque Waveform Determination Utilizing Speed Fluctuation," SAE Paper No. 890486.
- [6] Rizzoni, G., 1989, "Estimate of Indicated Torque From Crankshaft Speed Fluctuations: A Model for the Dynamics of the I.C. Engine," IEEE Trans. Veh. Technol., **38**(3), pp. 168–179.
- [7] Chen, K. S., and Chen, S., 1993, "Engine Diagnostics by Dynamic Shaft Measurement: A Progress Report," SAE Paper No. 932412.
- [8] Taraza, D., 1993, "Possibilities to Reconstruct Indicator Diagrams by Analysis of the Angular Motion of the Crankshaft," SAE Paper No. 932414.
- [9] Champoussin, J. C., and Ginoux, S., 1997, "Engine Torque Determination by Crank Angle Measurements: State of the Art, Future Prospects," SAE Paper No. 970532.
- [10] Guezennec, Y. G., and Gyan, Ph., 1999, "A Novel Approach to Real-Time Estimation of the Individual Cylinder Combustion Pressure for S.I. Engine Control," SAE Paper No. 1999-01-0209.
- [11] Gyan, Ph., Ginoux, St., Champoussin, J. C., and Guezennec, Y., 2000, "Crank-angle Based Torque Estimation: Mechanistic/Stochastic," SAE Paper No. 2000-01-0559.
- [12] Williams, J., 1996, "An Overview of Misfiring Cylinder Engine Diagnostic Techniques Based on Crankshaft Angular Velocity Measurements," SAE Paper No. 960039.
- [13] Taraza, D., Henein, N. A., and Bryzik, W., 1996, "Experimental Determination of the Instantaneous Frictional Torque in Multi-cylinder Engines," SAE Paper No. 962006.
- [14] Maass, H., and Klier, H., 1981, *Kraefte, Momente und deren Ausgleich in der Verbrennungskraftmaschine*, Springer, Wien.
- [15] Hafner, K. E., and Maass, H., 1985, *Torsionsschwingungen in der Verbrennungskraftmaschine*, Springer-Verlag, Wien.
- [16] Taraza, D., 1997, "A General Theory of V-Engines Balance as a Predictive Tool in Engine Design," *Proceedings of the Fall Technical Conference*, ASME, New York, ICE-Vol. 29-1, pp. 29–37.
- [17] Ventsel, H., 1973, *Théorie des Probabilités*, Editions Mir Moscou, Moscow.
- [18] Taraza, D., 2000, "A Probabilistic Approach to Engine Balance," ASME J. Eng. Gas Turbines Power, **122**, pp. 526–532.
- [19] Taraza, D., 2000, "Statistical Correlation Between the Crankshaft's Speed Variation and the Contribution of Individual Cylinders to the Total Engine Output," *Proceedings of the Fall Technical Conference*, ASME, New York, ICE-Vol. 35-3, pp. 81–93.

Statistical Correlation Between the Crankshaft's Speed Variation and Engine Performance—Part II: Detection of Deficient Cylinders and Mean Indicated Pressure Calculation

D. Taraza

Mechanical Engineering Department,
Wayne State University,
5050 Anthony Wayne Drive,
Detroit, MI 48202

The goal of this two-part paper is to develop a methodology using the variation of the measured crankshaft speed to calculate the mean indicated pressure (MIP) of a multicylinder engine and to detect cylinders that are lower contributors to the total engine output. The statistical model of a harmonic component of the engine torque developed in the first part of the paper is used to achieve this goal. The analysis of the half-order components of the gas pressure torque permits to identify distinct phase angle domains of the resultant torque vector that are specific for the deficiencies of given cylinders. Based on the rigid-body model of the crankshaft, these phase angle domains are correlated to the phase angle domains of the half-order component of the crankshaft speed. Then, the phase angle of the half-order component of the measured crankshaft speed will identify the deficient cylinder. The amplitude of the first major harmonic component of the measured crankshaft speed is correlated to the corresponding harmonic order of the gas pressure torque and is used to calculate the MIP of the engine. The accuracy limits of this "software dynamometer" are also presented. [DOI: 10.1115/1.1563245]

Introduction

In the first part of the paper, [1], it was shown that for the lower harmonic components of the cylinder torque the dynamic response of the crankshaft approaches the response of a rigid body. This situation permits a simple and direct correlation between the lower harmonic orders of the engine torque and of the crankshaft speed. The cycle variation of the torques contributed by each cylinder required the development of a statistical model for the harmonic components of the cylinder torque. It was found that the half-order component of the cylinder torque could be used to detect deficient cylinders and the first major harmonic order to estimate the MIP of the engine. Experimental investigation of a four-stroke six-cylinder automotive diesel engine is used to validate the theoretical model developed in Part I of the paper.

Experimental Investigation

Extensive experiments were conducted on a four-stroke six-cylinder direct injection diesel engine (Cummins 6CTA 8.3). The principal parameters of the engine are given in Appendix B at the end of the paper.

The engine was operated at constant speed and different loads, and several malfunctions have been simulated. To simulate a deficient cylinder, the nut connecting the high-pressure fuel line to the corresponding element of the injection pump was slightly unscrewed and a leakage was introduced in the fuel supply of the cylinder. By varying the tightness of the high-pressure fuel line, it

was possible to reduce gradually the amount of fuel injected into the cylinder from the rated value to zero, to simulate a wide range of malfunctions up to a complete misfire.

The pressures were measured in all cylinders by flush mounted piezoelectric pressure transducers (Kistler 6123) via charge amplifiers (Kistler 5010). The crankshaft's speed variation was determined using a high-precision hollow shaft encoder (BEI-5VL670 HAZ 10) with 360 divisions on the optical disk. The hollow encoder was mounted on the flywheel and permitted the coupling of the engine to the dynamometer. An absolute pressure transducer (OMEGA PX176-025 A5V) was used to measure the crankcase pressure and all experimental data were sampled and recorded by a 24-channel data acquisition system (DSP Technology) having an internal clock with a 10 MHz frequency.

The mean indicated pressure (MIP) and the gas pressure torque (GPT) of each cylinder were calculated from the pressure traces. The GPT and the measured speed were subjected to a Discrete Fourier Transform (DFT) to determine the amplitudes and phases of their harmonic components.

Simulation of a Misfiring Cylinder

For example, cylinder #4 was disabled by loosening the connection of the high-pressure fuel line to the corresponding element of the injection pump. Ten consecutive cycles were recorded and the MIP of each cylinder was calculated. Figure 1 shows the fluctuations of the MIP of each cylinder during this period. For the operating conditions corresponding to the situation shown in Fig. 1 one could assume that the MIP of cylinder #4 is 0.06 ± 0.02 MPa while the MIP of the rest of the cylinders is 0.84 ± 0.04 MPa. With these data the statistical parameters of the half-order component of the GPT are calculated according to the statistical model developed in Part I of the paper. The summary of the calculations is presented in Appendix A at the end of the paper and the results are shown in Fig. 2. The vector representing the half-order component

Contributed by the Internal Combustion Engine Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the ICE Division September 2001; final revision received by the ASME Headquarters July 2002. Associate Editor: D. Assanis.



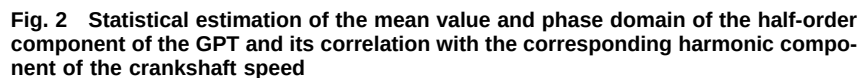
Simulating in the same manner different malfunctions of other cylinders, it was possible to determine the phase angle domains of the half-order component of the crankshaft speed, corresponding to the deficiency of each cylinder (Fig. 3).

While the phase angle of the half-order component of the measured crankshaft speed identifies a deficient cylinder, the amplitude of this component determines the degree by which the cylinder is a lower contributor to the total engine output.

If all cylinders are fairly equal contributors to the total engine torque, according to the statistical model developed in Part I of the paper, the amplitudes of the half-order components of the crankshaft speed will have a small value and their phase angles will be randomly scattered over a 360° domain (Fig. 4(a)). As a cylinder starts to contribute less to the total engine torque, the amplitude of the half-order component of the measured speed increases and the angular domain of its phase decreases and takes a preferential direction (Fig. 4(b, c)). When a cylinder completely misfires, the amplitude of the half-order component of the crankshaft's speed becomes very large and its phase varies very little from the preferential direction indicating the deficient cylinder (Fig. 4(d)).

The results obtained for six different malfunctions simulated on the experimental engine are represented in Fig. 5 for sequences of ten successive cycles.

Analysis of Fig. 5 shows that in the case of a misfire the phase angle of each individual cycle is placed in the expected domain as indicated by Fig. 3. This is true also for the case of cylinder #2 contributing about 35% less than the other cylinders. For the case of cylinder #6 contributing about 19% less than the rest of the cylinders the phases of individual cycles present a wider scatter. This situation is similar to the one shown in Fig. 4(b) and, for some cycles, cylinder #2 may be mistakenly considered as the deficient one. This ambiguity may be avoided if the phase of the harmonic order $k = 1\frac{1}{2}$ is also considered, or the half-order vector is averaged over a number of successive cycles, [2]. When the



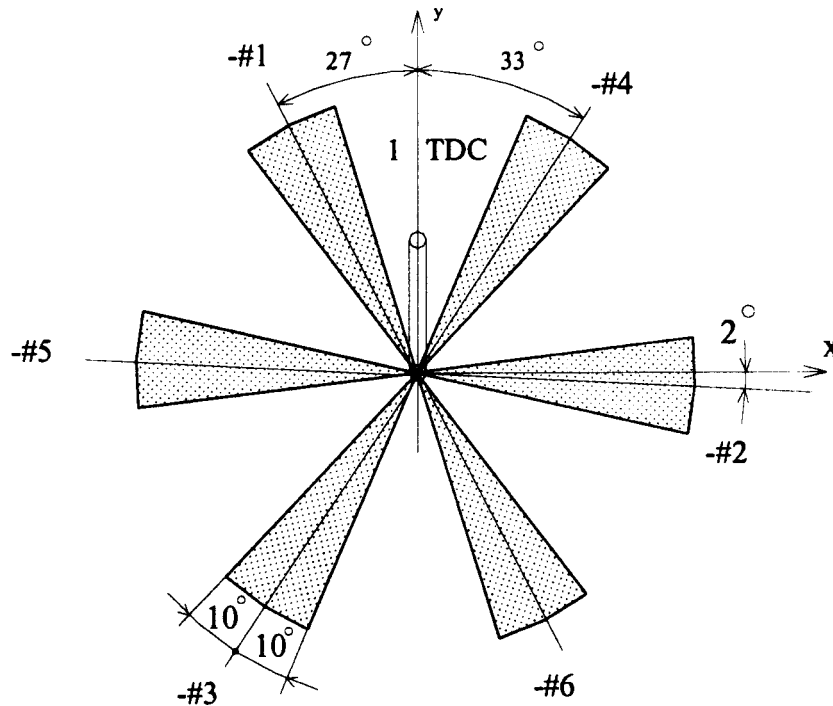


Fig. 3 The phase angles domains of the vectors representing the half-order component of the crankshaft's speed, corresponding to deficiencies of specific cylinders

amplitude of the half-order component of the crankshaft speed is less than 0.15 rad/sec the reduction in the contribution of the deficient cylinder is less than 20% and the responsible cylinder may be identified if the mean half-order speed vector is calculated for ten successive cycles (Fig. 6). The position and phase angle of the mean half-order vector speed could also identify cylinders that contribute about 7% to 10% less than the rest of the cylinders (Fig. 7).

The direction of the mean vectors in Fig. 6 and Fig. 7 are approximately the same, but the amplitudes of the vectors are different showing the different degree of deficiency of cylinder #6 in the two different cases.

The following limits of the mean value (over ten successive cycles) of the half-order speed vector were observed for the investigated engine.

- Fluctuations of $\pm 5\%$ with respect to the average MIP of the engine are normal for cycle-to-cycle and cylinder-to-cylinder

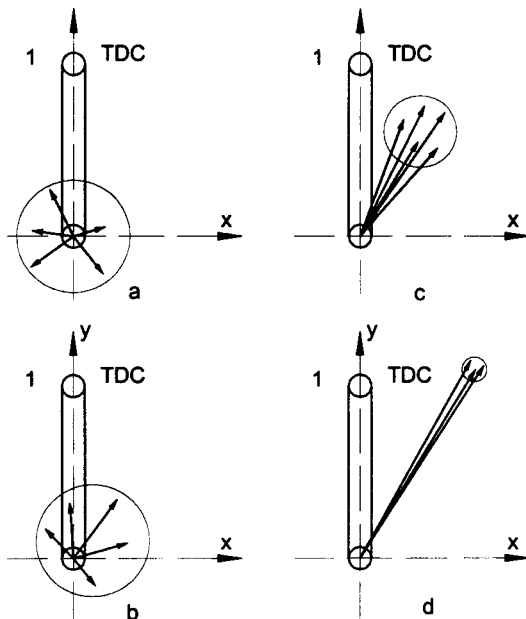


Fig. 4 Position of the ellipse of dispersion for the half-order component of the crankshaft speed, corresponding to different degrees of reduced contribution of cylinder #4

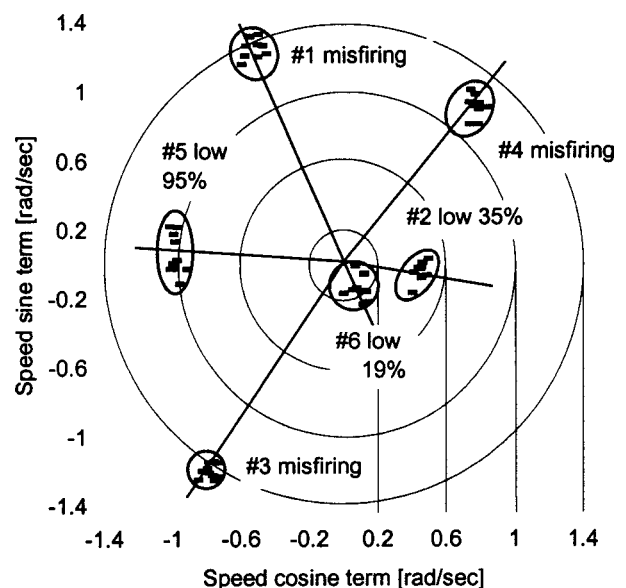


Fig. 5 Positions of the tips of the measured half-order speed vectors for different malfunctions in the operation of the cylinders

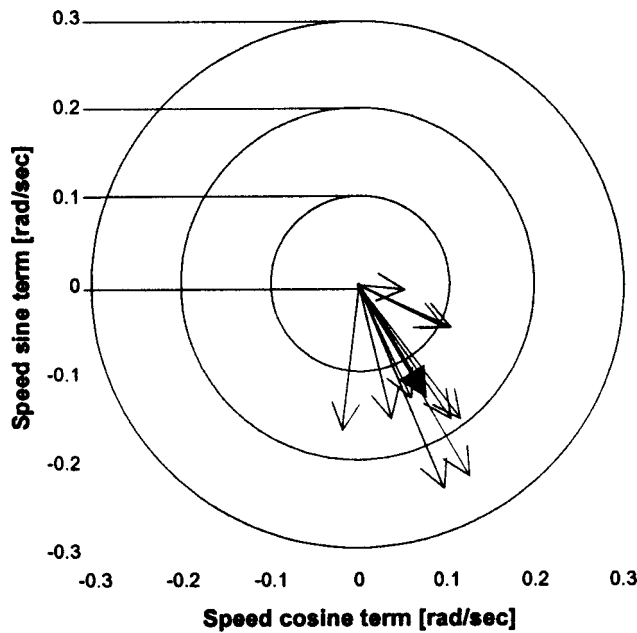


Fig. 6 The half-order vectors of the measured crankshaft's speed and their mean value (dark arrow) for cylinder #6 contributing 19% less than the other five cylinders

variation under steady-state operating conditions. In this case the mean value of the half-order speed vector is less than 0.08 rad/sec.

- Mean values included between 0.08 and 0.15 rad/sec correspond to a reduction of 7 to 20% in the contribution of the deficient cylinder. The faulty cylinder is identified by the phase angle of the half-order speed vector.
- Mean values included between 0.15 and 0.4 rad/sec correspond to a reduction of 20 to 35% in the contribution of the deficient cylinder.

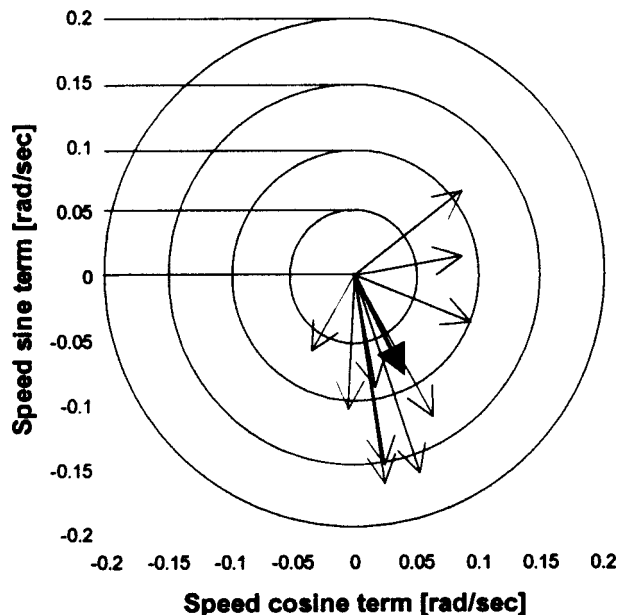


Fig. 7 The half-order vectors of the measured crankshaft's speed and their mean value (dark arrow) for cylinder #6 contributing 7% less than the other five cylinders

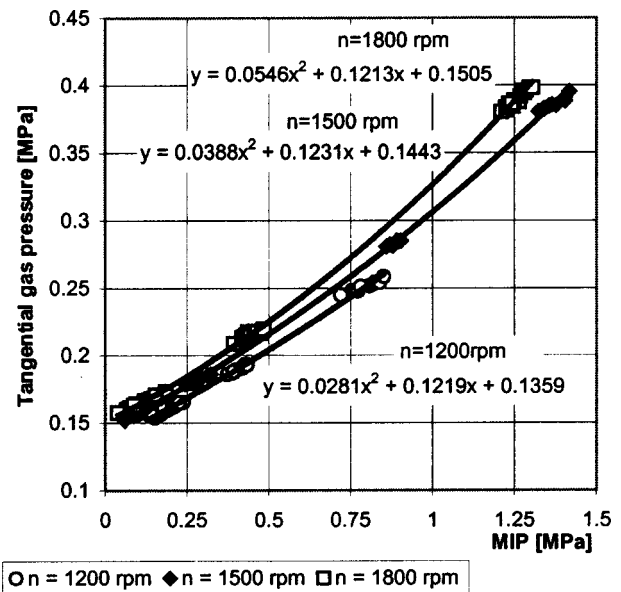


Fig. 8 Correlations between the MIP and the amplitudes of the third-order harmonic component of TGP for different engine speeds

- Mean values larger than 0.4 rad/sec indicate reductions of more than 35% in the contribution of the deficient cylinder and for complete misfires the mean values grow larger than 1.0 rad/sec.

These are typical values, but for a different engine more accurate limits may be obtained considering the statistical model of the half-order torque vector and the actual mass moment of inertia of the moving parts in Eq. (3) of Part I.

Calculation of the Average Mean Indicated Pressure (MIP) of the Engine

The correlation between the MIP and the amplitude of the third harmonic component of the tangential gas pressure shown in Figs. 6 and 7, of Part I of the paper may be used to estimate the MIP of the engine. More accurate results in the MIP calculations may be obtained if the influence of the engine speed is taken into consideration, [3]. Figure 8 presents the same data as Fig. 6 (pt. I) but separated as functions of speed.

If the crankshaft's speed is measured and subjected to a DFT, its third-order component $\dot{\theta}_3$ could be determined and represented in a system of coordinates having the cosine component on the horizontal axis and the sine component on the vertical axis (Fig. 9). The amplitude of the third-order GPT may be estimated from the correlations shown in Fig. 8 and its phase angle ϕ_3 from the correlation shown in Fig. 8 of Part I. The reciprocating inertia torque $T_{3_{ir}}$ is calculated from Eq. (31), of Part I, and represented parallel to the sine axis in Fig. 9. The phase angles are considered positive in clockwise direction and the torque vector phase is measured with respect to the positive sine-axis while the phase of the speed vector is measured with respect to the negative cosine-axis.

Because the third-order component of the crankshaft's speed should lag the torque by 90°, the third-order component of the resultant torque T_3 should be located at 90° clockwise with respect to the measured speed vector (Fig. 9). The amplitude of the third-order resultant torque T_3 may be estimated from Eq. (3) of Part I:

$$|T_3| = 3\omega J_{tot} |\dot{\theta}_3|.$$

Measurements have shown, [4,5], that the phase angle between the third-order speed vector $\dot{\theta}_3$ and the resultant of the gas pres-

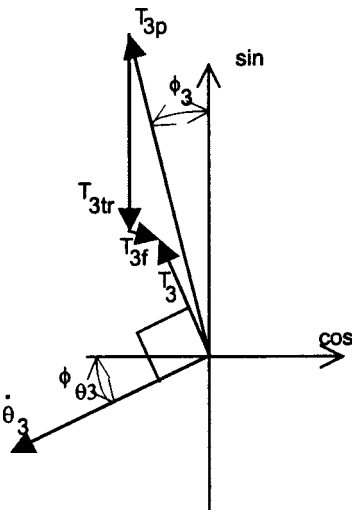


Fig. 9 Summation of the third-order torque vectors

sure and reciprocating inertia torques is close to 90°, but always a little smaller (Fig. 10). This result shows that the rigid model of the shafting is valid at low frequency excitation, but a small-third order torque component must be also considered. This residual component (T_{3f}) represents the contribution of friction and, possible, also a contribution of the first natural mode of the crankshaft that has been neglected in this analysis. This residual term may be also estimated from the vector summation in Fig. 9. Processing in this manner the data measured for different engine speeds and loads, the correlation shown in Fig. 11 was obtained.

Analysis of Fig. 11 shows that while the third-order components of the residual torque have small, but very different values at different speeds and loads, the third-order component of the resultant torque correlates almost linearly with the vector sum of the third order components of the gas pressure—and the reciprocating inertia—torques $|T_{3p} + T_{3tr}|$. Based on this correlation, the average value of the gas pressure torque may be obtained from the linear correlation shown in Fig. 11:

$$|T_{3p} + T_{3tr}| \cong 1.075|T_3| - 1.7456. \quad (1)$$

If the contribution of the residual torque is neglected, the calculations will slightly underestimate the mean value of the resultant torque.

The calculation of the third order component of the gas pressure torque T_{3p} from the third-order component of the measured crankshaft speed $\dot{\theta}_3$ may be made according to Fig. 9 following the steps shown in Table 1.

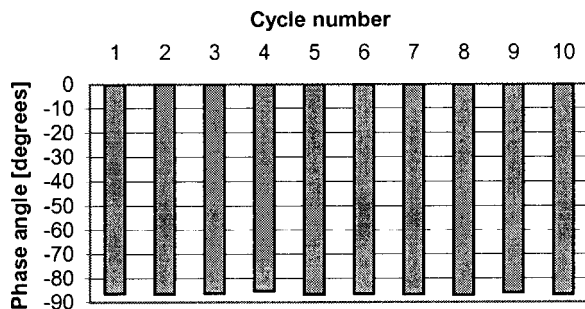


Fig. 10 The phase angle between the third-order speed vector and the vector representing the resultant of the GPT and RIT for ten successive cycles

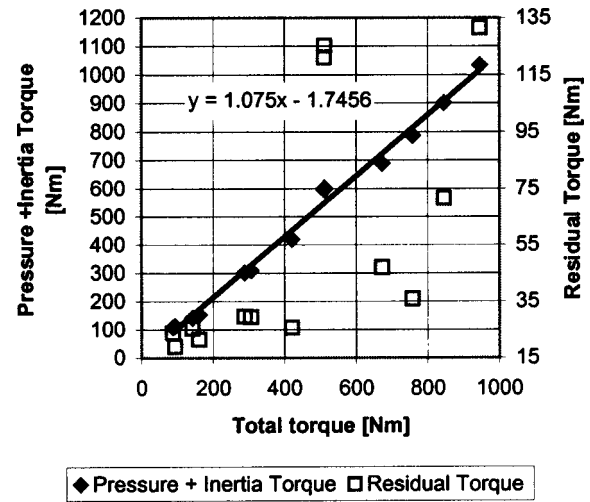


Fig. 11 Correlation between the third-order components of the torques acting on the crankshaft

- From the measured speed, the amplitude of the third harmonic component $|\dot{\theta}_3|$ and its phase ϕ_{θ_3} are obtained by the DFT of the measured data.
- The amplitude of the resultant third-order component of the engine torque $|T_3|$ is calculated by formula (3) of Part I and is phased 90° ahead (clockwise) of the speed vector.
- Expression (1) is used to estimate the contribution of the residual torque and to calculate the resultant of the gas pressure and reciprocating inertia- torques $|T_{3p} + T_{3tr}|$. Because all these vectors are subjected to a statistical scatter, the assumption is made that the phase angle of this resultant is the same as the phase angle of the total torque T_3 . This hypothesis implies that the third order component of the residual torque is small in comparison to the gas pressure torque and has an impact on the accuracy of the method at low engine loads.
- The resultant is then projected on the coordinate axes and the value of the reciprocating inertia torque calculated with formula (31) of Part I is added on the sine axis to obtain the sine component of the gas pressure torque (see Fig. 9).
- The amplitude of the gas pressure torque is calculated from

Table 1 Sample calculation of the MIP from the measured crankshaft's speed

Magnitude	Source	Value
ω	Measurement	188 rad/sec
$ \dot{\theta}_3 $	Measurement	0.8499 rad/s
ϕ_{θ_3}	Measurement	-21.56°
$ T_3 $	Equation (1.3) from Part I	757.4 Nm
$ T_{3p} + T_{3tr} $	Equation (1) from Part II	812.4 NM
T_{3ph}	$ T_{3p} + T_{3tr} \sin \phi_{\theta_3}$	-298.5 Nm
$ T_{3tr} $	Eq. (31) from Part I	834.9 Nm
T_{3pv}	$ T_{3p} + T_{3tr} \cos \phi_{\theta_3} + T_{3tr} $	1590 Nm
T_{3p}	$\frac{1}{6} \sqrt{T_{3ph}^2 + T_{3pv}^2}$	269.7 Nm
p_{tg_3}	Eq. (29) from Part I	0.391 MPa
MIP	Eq. (2) from Part II	1.270 MPa

its components on the sine and cosine axes and divided by the number of cylinders to obtain its mean value for one engine cylinder.

- Formula (29) of Part I is used to obtain the corresponding tangential gas pressure, which is further correlated with the MIP of the engine. In order to enhance the accuracy of the method, the more complex correlation function, considering the influence of the engine speed was developed, based on the measured data shown in Fig. 8:

$$\begin{aligned} \text{MIP} = & (33.058x^2 - 99.434x + 64.078)(p_{T_g})_3^2 \\ & + (-17.216x^2 + 50.886x - 26.371)(p_{T_g})_3 \\ & + (1.864x^2 - 5.5587x + 2.735) \end{aligned} \quad (2)$$

where x is the engine speed in rpm divided by 1000.

The measured data considered in Table 1 correspond to one engine cycle for which the average MIP, calculated from the pressures measured in the cylinders was 1.246 MPa. The calculations made in Table 1 lead to an error of 1.9% with respect to the actual value of the MIP in this cycle. These calculations have considered only mean values of the involved parameters.

If the residual torque is neglected, the calculations will yield a value of 1.226 MPa for the MIP underestimating the actual value by 1.6%. Neglecting the residual torque and replacing Eq. (2) by a simple correlation obtained from the generic data given in the literature, [6], a value of 1.331 MPa is found for the MIP, which overestimates the actual value by 5.2%. These results indicate the possibility to estimate the MIP relying only on generic data and, the order of magnitude of the error to be expected in this case.

The statistical model of a harmonic component of the gas pressure torque may be used to estimate the error limits in these calculations. In this case, it is necessary to consider a succession of cycles and determine the mean values and the standard deviations of the measured parameters. For example, if ten successive cycles are considered, at the same operating conditions as the cycle analyzed in Table 1, the following data are obtained:

$$\begin{aligned} \text{MIP} = & 1.2445^{+0.019}_{-0.016} \text{ MPa} \quad |\dot{\theta}_3| = 0.8463^{+0.0158}_{-0.0124} \text{ rad/sec} \\ n = & 1797.95^{+3.05}_{-2.75} \text{ rpm} \quad \phi_{\theta 3} = -21.74^\circ \pm 1.4^\circ_{-1.6^\circ} \end{aligned}$$

From these limits of variation, the following standard deviations were estimated:

$$\begin{aligned} \sigma_p \cong & 0.006 \text{ MPa}, \quad \sigma_{\dot{\theta}} \cong 0.005 \text{ rad/sec}, \quad \sigma_n \cong 0.96 \text{ rpm}, \\ \sigma_{\phi_{\theta}} \cong & 0.05^\circ = 8.78 \times 10^{-4} \text{ rad}. \end{aligned}$$

By determining the statistical parameters for the succession of algebraic operations shown in Table 1 it is found, for the calculated MIP, a standard deviation of 0.0067 MPa. Considering limits of $\pm 3\sigma$ for the predicted MIP, the result of the calculations may be expressed as

$$\text{MIP} = 1.27 \pm 0.02 \text{ Mpa.}$$

While the limits of variation are very close to the measured ones, the mean value is shifted upward 2.14% with respect to the measured mean value. Sample results of such calculations, made for successions of ten engine cycles at different steady-state operating conditions are given in Table 2.

According to these results, the errors affecting the calculation of the MIP from the crankshaft's measured speed may reach a maximum of 5% in the upper and 7% in the medium range of the engine load. At loads corresponding to an MIP of less than 0.4 MPa the errors may reach 30% and the method is no more reliable. The source of errors lays mainly in the impossibility to correctly estimate the third-order component of the residual torque. At higher loads the residual torque is a small fraction of the gas pressure torque and the errors are small. At lower engine loads the residual torque becomes comparable to the gas pressure torque and considerable increases the errors.

Table 2 Error estimate in the calculations of the MIP from the measured crankshaft speed

$ \dot{\theta}_3 $ Rad/sec	$\phi_{\theta 3}$ Degrees	Speed rpm	MIP (MPa) Calculated/ Measured	Error %
1.2632	-11.13	1506.9	1.327/1.357	-2.2
1.0193	-17.85	1682.4	1.3022/1.3448	-3.17
0.8462	-21.74	1797.9	1.271/1.2445	2.14
1.0948	-18.29	1364.4	1.0452/1.1018	-5.1
0.6902	-8.61	1497.7	0.857/0.8794	-2.54
1.1588	-6.81	1187	0.8354/0.8217	1.67
0.3567	-15.58	1649.7	0.6599/0.6203	6.38
0.0756	-1.94	1792.4	0.4475/0.4491	0.1
0.7063	-1.94	1197.5	0.4136/0.4036	2.47
0.121	-12.96	1681.9	0.3929/0.366	7.23
0.166	-78.78	1497.5	0.2027/0.289	-29.9
0.5098	-0.03	1198	0.199/0.199	0
0.1812	165.98	1797.9	0.0976/0.1115	-12.5

Summary and Conclusions

At steady-state operating conditions, the measured crankshaft speed of a diesel engine carries the necessary information for determining the average gas pressure torque (or MIP) of the engine and for detecting malfunctioning cylinders. Subjecting the speed to a DFT and considering only its lower harmonic components may avoid the distortion produced by torsional vibrations on the measured speed.

The half-harmonic component of the crankshaft speed identifies cylinders that are lower contributors to the total engine output. If the half-order components have low amplitude (less than 0.08 rad/sec) and their phase angle is arbitrarily scattered over 360°, the differences between the contributions of the cylinders to the total engine torque are lower than 5%. If the contribution of a cylinder is reduced by 7 to 20% with respect to the others, the amplitude of the half-order component of the measured speed increases from 0.08 to 0.15 rad/sec. A complete misfiring of a cylinder is characterized by amplitudes larger than 1 rad/sec. These limits may be slightly different for different engines depending mainly on the total mass moment of inertia of the crankshaft.

While the amplitude of the half-order component of the measured crankshaft speed determines the degree of deficiency of a cylinder, its phase angle identifies the responsible cylinder. The statistical model of the half-order component of the crankshaft speed determines the phase angle domains that are characteristic for the deficiency of each cylinder of the engine. Cylinders with contributions reduced by only 6–7% may be identified by the phase angle of the half-order component, averaged over ten cycles.

The amplitude of the first major harmonic component of the crankshaft speed may be used to estimate the MIP and gas pressure torque of the engine. The corresponding harmonic component of the residual torque cannot be determined, but it was found that the sum of the gas pressure and reciprocating inertia torques could be linearly correlated to the total engine torque. Using this correlation and the correlation between the MIP and the amplitude of the first major harmonic order of the crankshaft speed makes it possible to estimate the MIP of the engine.

The error affecting the MIP calculated by this method is about 5% in the upper range, and increases to about 7% in the middle range of the engine power. At very low engine output, the error could increase up to 30%. The errors may be slightly higher if the residual torque is neglected and the correlation between the MIP and the TGP is based on generic data given in the literature.

The calculations required to determine the MIP and to identify malfunctioning cylinders are fairly simple and can be implemented in the engine controls to monitor in real time the healthy operation of the engine.

Table 3 Nonuniform cylinder operation (cylinder #4 disabled), 1200 rpm and MIP=0.84±0.04 MPa. Summary of calculations (half harmonic order $k=1/2$).

Parameter	Source	Value
Average MIP (N-1) cylinders	Assumed	0.84±0.04 MPa
MIP (deficient cylinder)	Assumed	0.06±0.02
$(p_{T_{G_{0.5}}})_{1,2,3,5,6}$	Fig. 6 (Part I)	0.31±0.02 MPa
$(\varphi_{0.5})_1$	Fig. 8 (Part I)	-133°±0.5°
$(p_{T_{G_{0.5}}})_4$	Fig. 6 (Part I)	0.105±0.005 MPa
$(\varphi_{0.5})_4$	Fig. 8 (Part I)	-165°±2°
$(T_{0.5})_{1,2,3,5,6}$	Equation (29)	213.6±13.8 Nm
$(T_{0.5})_4$	Equation (29)	72.3±3.4 Nm
$(\sigma_{T_{0.5}})_{N-1}$	One third of the limits	4.6 Nm
$(\sigma_{\varphi_{0.5}})_{N-1}$	limits	0.0029 rad
$(\sigma_{T_{0.5}})_4$	of variation	1.15 Nm
$(\sigma_{\varphi_{0.5}})_4$		0.0116 rad
$(\bar{T}_{0.5})_{N-1}$	Equation (15)	204.2 Nm
$(\bar{T}_{0.5})_{N-1}$	Equation (15)	-62.4 Nm
$(\bar{T}_{0.5})_4$	Equation (4)	-69.8 Nm
$(\bar{T}_{0.5})_4$	Equation (4)	-18.7 Nm
$\bar{T}_{0.5x}$	Equation (20)	134.4 Nm
$\bar{T}_{0.5y}$	Equation (20)	-81.1 Nm
$\bar{T}_{0.5}$	$\bar{T}_{0.5} = \sqrt{\bar{T}_{0.5x}^2 + \bar{T}_{0.5y}^2}$	157 Nm
$(\sigma_{T_{0.5x}})_{N-1}$	Equation (16)	6.7 Nm
$(\sigma_{T_{0.5y}})_{N-1}$	Equation (16)	7.9 Nm
$(\sigma_{T_{0.5x}})_4$	Equation (5)	0.86 Nm
$(\sigma_{T_{0.5y}})_4$	Equation (5)	1.13 Nm
$\sigma_{T_{0.5x}}$	Equation (21)	6.75 Nm
$\sigma_{T_{0.5y}}$	Equation (21)	7.98 Nm
$(K_{0.5xy})_{N-1}$	Equation (17)	5.8 (Nm) ²
$K_{x_4y_4}$	Equation (7)	0.15 (Nm) ²
$K_{0.5xy}$	Equation (22)	5.95 (Nm) ²
$\beta_{0.5}$	Equation (23)	16.66°
σ_{ξ}	Equation (24)	7.09 Nm
σ_{μ}	Equation (24)	7.67 Nm

All the equations used in these calculations are taken from Part I of the paper and refer to the statistical model of a harmonic component of the GPT.

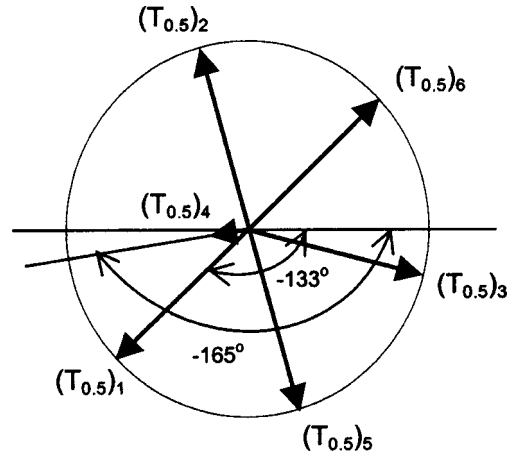


Fig. 12 The phase angle diagram of the half-order torque vectors; cylinder #4 disabled

The phase angle diagram of the half-order torque vectors considered in the calculation is shown in Fig. 12.

Appendix B

The investigated engine has the following geometric and mass parameters:

- Bore $D=114$ mm
- Stroke $S=135$ mm
- Crank radius $R=S/2$
- Connecting rod length $L=216$ mm
- Reciprocating mass $m_{tr}=3.557$ kg
- Total polar mass moment of inertia of the shafting $J_{tot}=1.58$ kg·m²

References

- [1] Taraza, D., 2003, "Statistical Correlation Between the Crankshaft's Speed Variation and Engine Performance—Part I: Theoretical Model," *ASME J. Eng. Gas Turbines Power*, **125**, pp. 791–796.
- [2] Taraza, D., 2001, "Statistical Model and Simulation of Engine Torque and Speed Correlation," *SAE Paper No. 2001-01-3686*.
- [3] Taraza, D., 2002, "Accuracy Limits of IMEP Determination from Crankshaft Speed Measurements," *SAE Paper No. 2002-01-0331*.
- [4] Taraza, D., 2000, "Statistical Correlation Between the Crankshaft's Speed Variation and the Contribution of Individual Cylinders to the Total Engine Output," *Proceedings of the Fall Technical Conference ASME*, New York, ICE-Vol. 35, pp. 81–93.
- [5] Taraza, D., 2001, "Quantifying Relationships Between the Crankshaft's Speed Variation and the Gas Pressure Torque," *SI Engine Modeling and Simulation (SP-1606)*, SAE, Warrendale, PA, pp. 89–101.
- [6] Hafner, K. E., and Maass, H., 1985, *Torsionsschwingungen in der Verbrennungskraftmaschine* Springer-Verlag, New York, pp. 141–143.

S. L. Yang
Associate Professor

Y. K. Siow

B. D. Peschke

Department of Mechanical Engineering and
Engineering Mechanics,
Michigan Technological University,
1400 Townsend Drive,
Houghton, MI 49931

R. R. Tacina

Aerospace Engineer,
Combustion Technology Branch,
NASA John H. Glenn Research Center,
Lewis Field,
21000 Brookpark Road,
Cleveland, OH 44135

Numerical Study of Nonreacting Gas Turbine Combustor Swirl Flow Using Reynolds Stress Model

This paper presents recent research on the use of a Reynolds stress turbulence model (RSTM) for three-dimensional flowfield simulation inside gas turbine combustors. It intends to show the motivations for using the RSTM in engine flow simulation, to present a further validation of the RSTM implementation in the KIVA code using the available experimental data, and to provide comparisons between RSTM and $k-\epsilon$ turbulence model results for chemically nonreacting swirling flows. The results show that, for high-degree swirl flow, the RSTM can provide predictions in favorable agreement with the experimental data, and that the RSTM predicts recirculations and high velocity gradients better than does the $k-\epsilon$ turbulence model. The results also indicate that the choice of swirler has a significant influence on the structure of the combustor flowfield.

[DOI: 10.1115/1.1560706]

Introduction

A key characteristic of gas turbine combustor performance is the emission of NO_x . Although the emissions of unburned hydrocarbons and carbon monoxide have been greatly reduced through the design process, the problem of NO_x emission still requires significant improvements in combustor design, [1].

Combustion at near stoichiometric fuel-air ratios produces high combustion temperatures which are the main cause of high thermal NO_x emissions. As a result, combustor designs have sought nonstoichiometric fuel-air ratios in both rich and lean ranges. Rich burn combustion reduces NO_x by limiting the amount of oxygen present for high temperature combustion. Combustors like these are often labeled rich-burn quick-quench lean-burn (RQL). Here, a rich burn primary zone is followed by a quick quench zone where the combustion gases are diluted before combustion is completed in a cooler lean burn zone. This design concept has been the subject of several studies, [2–5]. An alternative to RQL is lean-premixed combustion in which a lean combination of fuel and air is premixed and prevaporized prior to introduction into the combustor. It has also been studied by several researchers, [6–8], and it was found that low NO_x emissions required uniform fuel-air mixtures. Tacina [9] found that a lean premixed combustor produced less NO_x in comparison with the RQL.

A third approach is called lean direct injection (LDI). Instead of injecting a lean premixed quantity of fuel and air into the combustor, fuel is injected into the airflow entering the combustor. Such a combustor depends on fuel injectors which can produce a spray of very fine droplets that quickly atomize and vaporize. Furthermore, the mixing process is also a key issue for the LDI.

Several studies have pursued computational fluid dynamics techniques to predict the flowfield inside the combustor. Successful predictions of turbulent reactive flows depend highly on successful prediction of the turbulent fluid flow, [10]. A popular choice for the turbulence model has been the two-equation $k-\epsilon$ eddy viscosity turbulence model (EVTM). Such calculations have been performed, [3,11–13]. Despite its popularity and low computational overhead, the standard $k-\epsilon$ model suffers from several

deficiencies. It is known for being inappropriate for modeling swirling flows, [14], which are often present and desired in combustors and IC engines. More often than not, ad hoc modifications to the $k-\epsilon$ equations are necessary to provide satisfactory results. It also tends to incorrectly predict recirculation regions, [10]. Prediction of these flow structures is important since the location of recirculation regions can influence the residence time involved in NO_x production, [5], and/or can increase the residence time for liquid fuel evaporation. Finally, the standard $k-\epsilon$ model often underpredicts large velocity gradients because of the diffusive nature of this model, [13].

In contrast, the Reynolds stress turbulence model (RSTM) provides hope for better accuracy in predicting the complex swirling flow, which may also contain recirculation. Unlike the standard EVTm, the RSTM is able to capture the anisotropy caused by complex flow, flow rotation (swirl flow), streamline curvature, wall effects, etc. This is achieved *naturally* since each Re-stress component has its own transport equation. As a result, the RSTM is needed to properly address these flow issues.

Re-stress calculations have been performed on combustors, [15,16], for cases with low and high inlet swirl. In each study the Gibson and Launder [17] RSTM was used with a modified turbulence specific dissipation (ϵ) equation. Meanwhile, promising results were observed in an RSTM study by Yang et al. [18] when applying the Speziale-Sarkar-Gatski (SSG) model of Speziale et al. [19] to reciprocating engines and to a lean premixed combustor.

The purpose of this study is to expound upon Reference [18] by further investigating and validating the SSG RSTM for nonreacting swirl flow calculation of an LDI combustor using the KIVA-3V computer code, [20]. The study is to substantiate the RSTM as a preferred turbulence model by presenting evaluations and comparisons of the results with available experimental data and $k-\epsilon$ turbulence model results. It is also desired to show that, unlike two-equation models, a RSTM does not require ad hoc modification, while being capable of capturing the complex characteristics of turbulent swirling flow.

Reynolds Stress Turbulence Model (RSTM) and Code Implementation

In Ref. [18], the performance of the SSG RSTM for simulating turbulent flow inside the cylinder of reciprocating engines and the

Contributed by the Internal Combustion Engine Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the ICE Division, February 2002; final revision received by the ASME Headquarters, March 2002. Editor: D. N. Assanis.

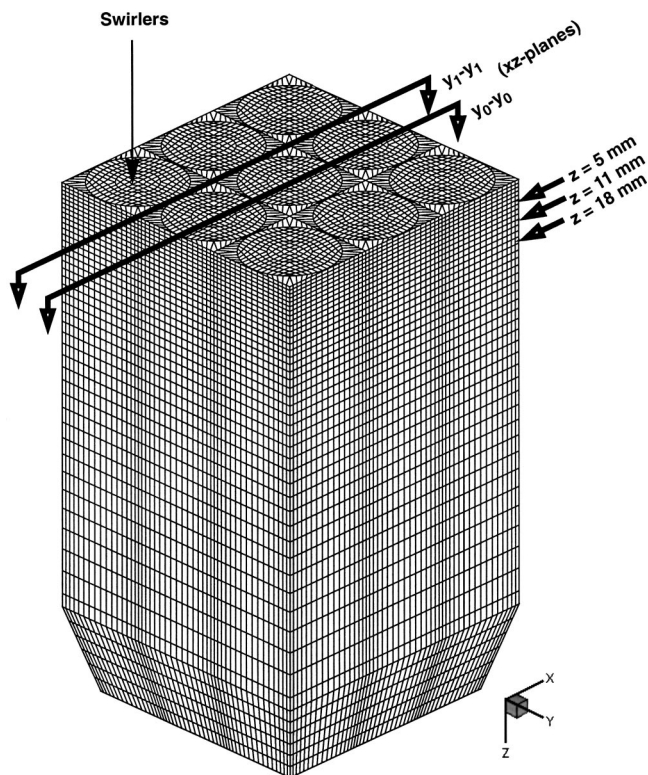


Fig. 1 Combustor mesh and cutting planes

combustor of a gas turbine was reported. One major advantage of using the SSG model is that, unlike most other RSTMs available in the literature, no wall-distance parameter is needed. Because of the complex geometry involved, this feature is especially desirable and important for engine flow simulation. Due to space limitation, the SSG modeling equations, details of the KIVA code implementation, and code validation will not be repeated here, but can be found in Ref. [18]. In this paper, additional model validation and performance of the SSG model will be given below.

Description of the Problem

For a gas turbine combustor, swirl flow is used to improve combustor performance by aiding in the fuel-air mixing process and by producing recirculation regions which can act as flame holders for flame stability. As mentioned earlier, recirculation can influence residence time and, as a result, NO_x production. Therefore, proper selection of a swirler is needed to reduce NO_x .

In this study, an LDI combustor is considered. The modeled LDI combustor includes inlets at nine inflow locations, which are attached to a single primary combustion zone shown in Fig. 1. The inlets are atop helical axial swirlers arranged in a three-by-three square which lies in the xy -plane. The rotation imparted to the air by each swirler is directed in the axial or z -direction. Two different swirlers were used based on blade angle inclination from the direction of airflow. One swirler had a 45 deg angle, and the other 60 deg. These angles refer to the outermost vane twisting angle of the helical swirler. A plot of a 60 deg swirler attached to a convergent-divergent venturi is shown in Fig. 2. The central hol-

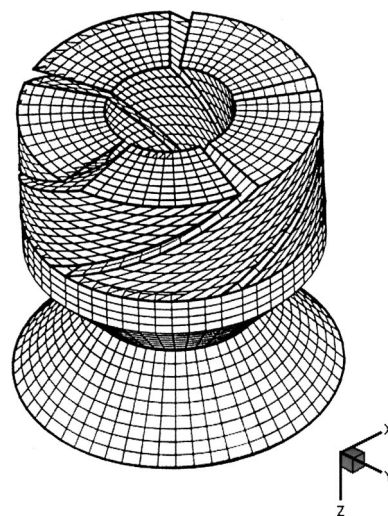


Fig. 2 A sketch of the swirler

low region is reserved for the fuel injector while the five helical hollow slots represent the blades that guide the airflow.

The computational domain begins with the inlets placed at the top of nine 45 deg or 60 deg swirlers. These swirlers are either co-rotating clockwise, or counter-rotating. After air exits a swirler it passes through a venturi. At the center of the venturi circle, a fuel injector nozzle can inject liquid fuel. This arrangement is used since the combination of swirler and venturi were found to have a significant influence on the liquid atomization and droplet distribution, [21]. A short divergent cone, acting like a diffuser, follows and ends at the primary zone inlet plane. Two constant y -planes and three constant z -planes were defined for postprocessing, where z represents the axial direction. These cutting plane locations are shown in Fig. 1.

Notable boundary conditions include the wall and inlet boundaries. Since the turbulence equations presented are valid only for high-Re number flows, wall functions were used for wall boundaries. Diffusion fluxes were set to zero at the walls. The specified inlet velocity value depends on the chosen operating conditions of the combustors. These conditions include a pressure drop, Δp , of 3% across the height of the combustor, an outlet pressure of 1.0 atm, and an inlet temperature of 27 degC. These quantities were then used to determine the inlet air velocity V_{in} , which is uniform at all nine swirlers, according to the following equation:

$$V_{in} = C_D \sqrt{2(\Delta p)/\rho}. \quad (1)$$

The discharge coefficient C_D was experimentally determined and provided by the sponsor. Uniform inlet turbulence quantities of k_{in} , ε_{in} , and Reynolds stresses are listed in Table 1.

Three computational meshes with different grid densities, namely 113793, 173829, and 254853 grid points were used. Results from the three grids are consistent with each other. The medium density grid was used in the following presentations to represent the results of all three grids.

Results and Discussions

The purpose of this study was to investigate the nonreacting swirl flow within the combustor. Therefore, the spray and ignition

Table 1 Inlet turbulence values

$\bar{u}_{in}$	k_{in}	ε_{in}	$\tilde{u}_1 \tilde{u}_1$	$\tilde{u}_2 \tilde{u}_2$	$\tilde{u}_3 \tilde{u}_3$	$\tilde{u}_1 \tilde{u}_2$	$\tilde{u}_1 \tilde{u}_3$	$\tilde{u}_2 \tilde{u}_3$
0.05 V_{in}	1.5 $\bar{u}_{in}^2$	0.25 d	0.25 k_{in}	0.65 k_{in}	1.1 k_{in}	0.0	0.0	0.0

options were turned off. Calculations were performed with the SSG RSTM and the $k-\varepsilon$ model using the quasi-second-order upwind scheme in the KIVA code. The computed results were compared and validated with the experimental data for the 60 deg co-rotating case. In addition, a 45 deg co-rotating and a 60 deg counter-rotating-at-center case were also investigated. Since the computational domain was three-dimensional, results of these predictions are presented at the cutting planes indicated in Fig. 1.

Experiment. The experiments were performed at the NASA Glenn Research Center (GRC) CE-5. A square duct of quartz walls was used as the primary combustion zone for optical access. The sides of the square were 76.2 mm wide. A phase/Doppler particle analyzer was used for all velocity measurements. A two-component system using green (514.5 nm) and blue (488.0 nm) beams from an argon-ion laser operating at 1.5 W power output was used for the measurements. The transmitting optics utilized a 500-mm focal length lens combined with a 300-mm focal length collimating lens to yield a focused beam waist of $131\ \mu\text{m}$ for the green and $124\ \mu\text{m}$ for the blue lines. The fringe spacing was $6.788\ \mu\text{m}$ for the green and $6.667\ \mu\text{m}$ for the blue lines. The receiving optics were located 30 degrees off-axis in the forward-scatter direction. Light was collected using a 500-mm focal-length f5.4 lens and then focused onto a $100\text{-}\mu\text{m}$ by 1-mm long slit. Further details of the instrument can be found in Ref. [22], and details of the experimental procedure can be found in Ref. [23]. The experimental results are presented in Figs. 3(c)–5(c) and are discussed below.

60 Deg Corotating Swirlers Case. Observations of the results on the xy -planes show that the RSTM compares more closely to the experiments than does the $k-\varepsilon$ model. The plots in Figs. 3–5 display computational results across xy -planes, at $z = 5\text{ mm}$, 11 mm , and 18 mm , respectively, while the experimental plots show the data measurements at the upper right corner of the same xy -planes. A coordinate system is displayed on the figures to indicate the x and y -directions. The positive z -direction, from the swirler inlet to the combustor outlet, points into the page.

Figure 3 paints a distinct picture that demonstrates the superiority of the RSTM when comparing its result with the $k-\varepsilon$ and experimental plots. Between the two turbulence models, the magnitude of the velocities differ greatly, although the axial velocity (w) contour structures and the vector orientations display similarities. Comparison with experimental data in Fig. 3 shows that the reversed flow at the center of the circles is better captured by the RSTM. The contour legend shows that the RSTM velocity levels vary over a greater range as seen in the experimental data plot.

At $z = 11\text{ mm}$ in Fig. 4, the RSTM continues to exhibit its capability as validated by the experiment. Unlike the $k-\varepsilon$ model result where no recirculation can be observed, the RSTM plot clearly shows a reversed flow region at each swirler center. Similar observations are found further downstream at the $z = 18\text{ mm}$ surface in Fig. 5. The RSTM result shows that the forward flow between swirlers and reversed velocity magnitude at the swirler centers are in good agreement with the experiment. This shows that the RSTM can better predict the reversed flow associated with recirculations. It also shows its ability to resolve the effects of strong swirl as verified by the experimental data.

Additional observations can be made from Figs. 3–5. The four subregions enveloped by any four neighboring circles contain negative axial velocities (i.e., reversed flow) in the RSTM results, whereas the $k-\varepsilon$ plots present little (Figs. 3, 4) and no (Fig. 5) reversed flow in these regions. Furthermore, these four subareas contain swirls in the counterclockwise (ccw) direction, opposite to the primary flow direction from the nine swirlers. This is due to the opposing velocities surrounding the central circle, where the shear force in one direction impedes and steers away the shear force in the other, creating the four counterclockwise swirls.

The contrast between the two models can be further displayed by viewing the velocity on the xz -planes, such as along the $y_0\text{-}y_0$

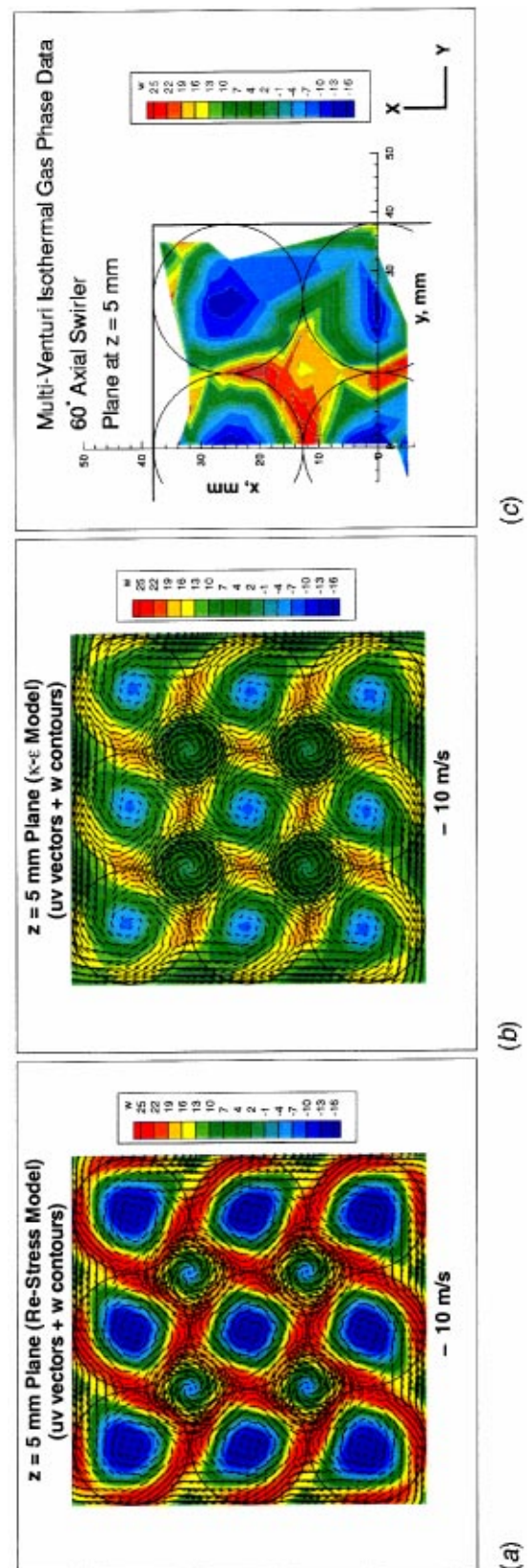


Fig. 3 Results for 60 deg swirler at $z = 5\text{ mm}$ cutting plane: (a) Reynolds stress turbulence model, (b) $k-\varepsilon$, (c) experiment

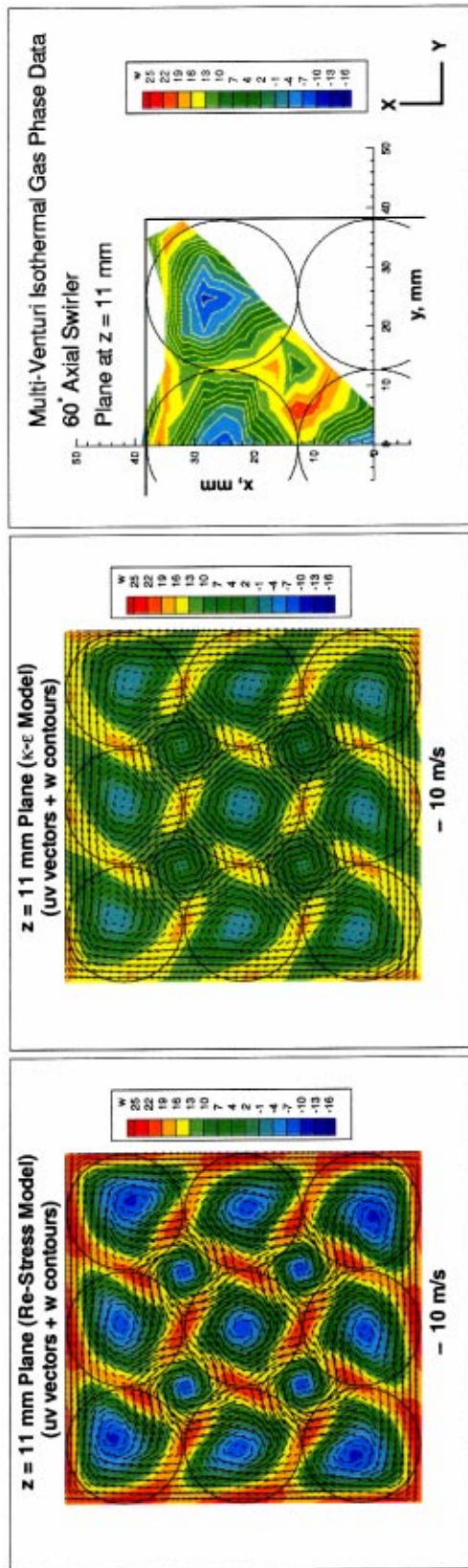


Fig. 4 Results for 60 deg swirler at $z=11$ mm cutting plane: (a) Reynolds stress turbulence model, (b) $k-\epsilon$, (c) experiment

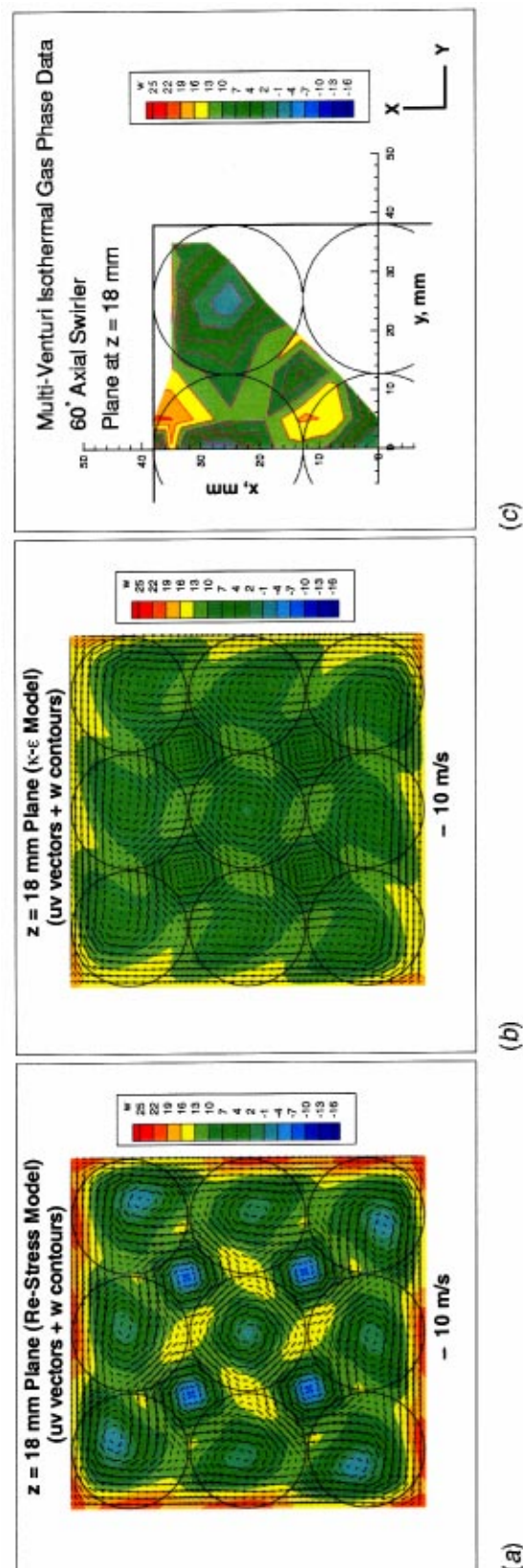
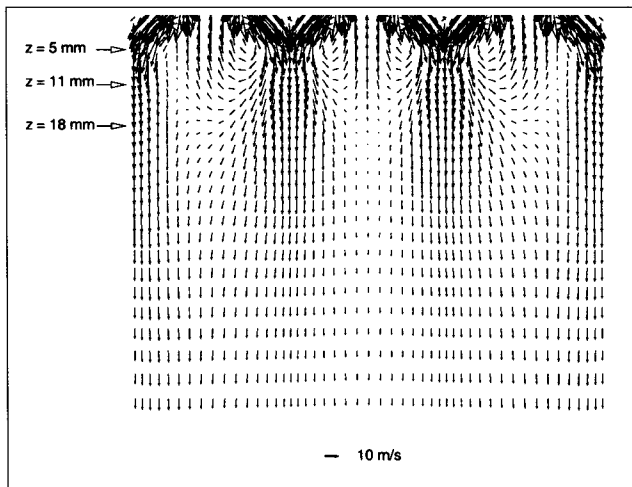
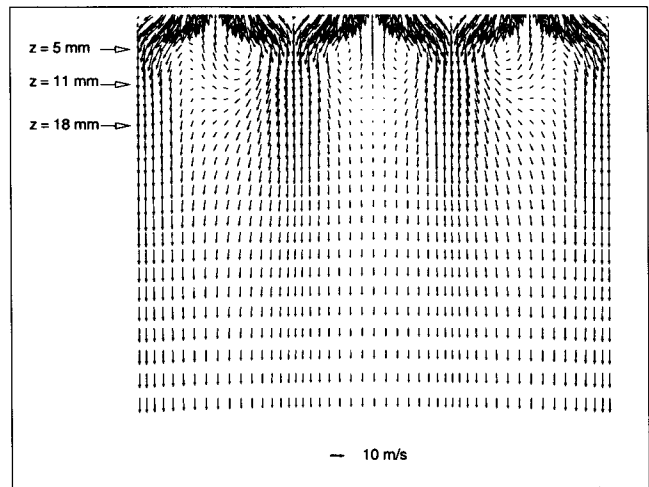


Fig. 5 Results for 60 deg swirler at $z=18$ mm cutting plane: (a) Reynolds stress turbulence model, (b) $k-\epsilon$, (c) experiment



(a) RSTM



(b) $k-\epsilon$

Fig. 6 Velocity on y_0 - y_0 plane for 60 deg swirler: (a) Reynolds stress turbulence model, (b) $k-\epsilon$

cutting plane shown in Fig. 6. Reversed flow is much stronger below the swirlers with the RSTM. For example, the RSTM reverse flow reaches up to the venturi while the $k-\epsilon$ result falls far short. Concurrently, the forward flow regions predicted by the RSTM are also stronger. Since much of the cross section of the combustor has a reversed mass flux, the forward mass flux regions must be stronger to satisfy overall mass conservation.

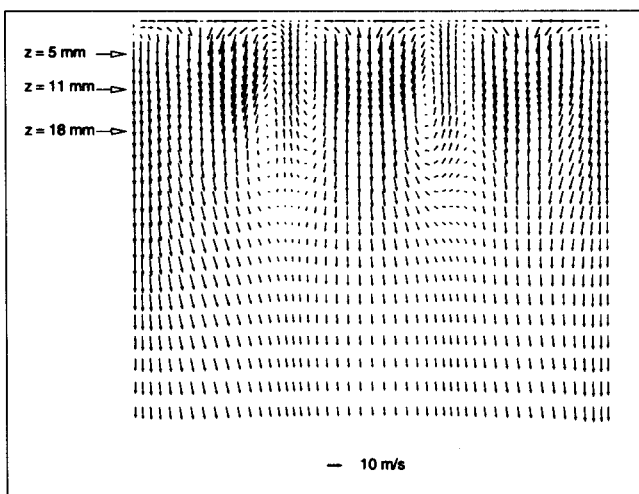
Differences between the turbulence models can also be shown on the y_1 - y_1 cutting planes, which do not cut through the swirlers. According to Fig. 7, reversed flow exists on this plane. The RSTM result shows two deep recirculation regions on the plane, while the $k-\epsilon$ model in Fig. 7 does not predict any reversed flow at all.

The $k-\epsilon$ model results appear more diffusive than those of the RSTM. This is due to the overestimated eddy viscosity which is a known characteristic of the standard $k-\epsilon$ model. This set of results also indicates that without ad hoc modification, the $k-\epsilon$ model is simply unable to capture the essence of flow recirculation from locally large axial pressure gradients, $\partial p / \partial z$, created by the high degree of swirl, [24]. As air exits from a swirler into the combustor, its swirling motion creates strong pressure gradients in the radial direction, a phenomenon similar to a free swirling jet being ejected from a nozzle. As the flow carries itself further down-

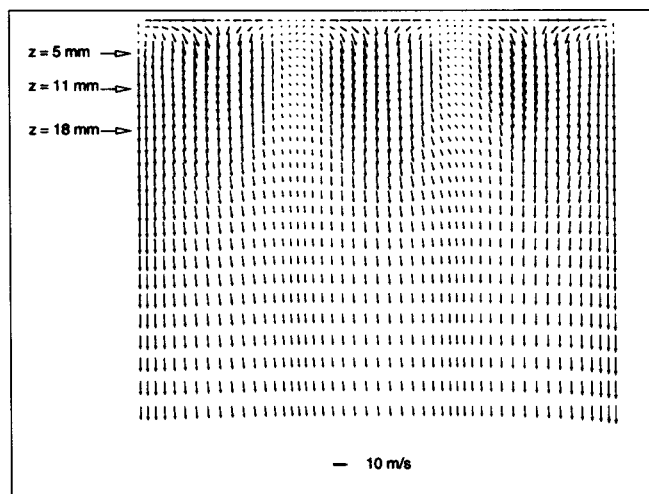
stream, an adverse pressure gradient builds up within the boundary of the air stream, causing the flow to reverse its axial direction and hence, a recirculation zone is created downstream of each swirler.

In short, with the RSTM whose Re-stress components are each computed with its own transport equation, the swirling and complex flow field can be resolved to an extent agreeable to the experiment. It can therefore be concluded that the RSTM has been correctly implemented, and that it is a valid model for the additional result comparisons discussed below.

45 Deg Swirlers Case Study. The results obtained with the 45 deg swirler primarily display the significant difference between swirlers. The magnitude of the inlet velocity is the same as in the 60 deg swirler case. As Figs. 8 and 9 indicate, there is no reversed flow at the center of the circles. Since the swirler diverts the upstream air at a smaller angle, a strong and solid forward-flow jet enters the combustor and does not permit recirculation as did the more conical jet from the 60 deg swirlers. Due to the lower degree of swirl, both the turbulence models produce results almost identical to each other; however, weak recirculation regions are cap-



(a) RSTM



(b) $k-\epsilon$

Fig. 7 Velocity on y_1 - y_1 plane for 60 deg swirler: (a) Reynolds stress turbulence model, (b) $k-\epsilon$

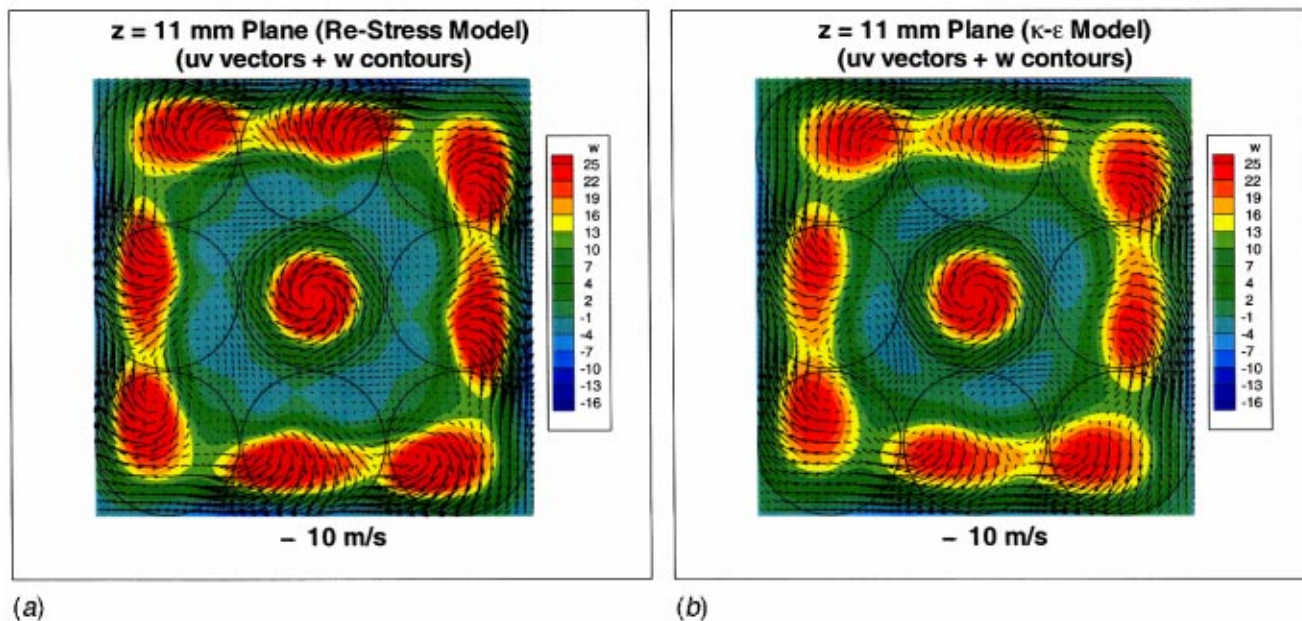


Fig. 8 Results for 45 deg swirler at $z=11$ mm plane: (a) Reynolds stress turbulence model, (b) $k-\epsilon$

tured by the RSTM. Since the 45 deg swirler produces such weak swirl flow, these results show that both RSTM and $k-\epsilon$ models can predict the flow field equally well.

The velocity profiles in Fig. 9 suggest that the atomization, vaporization, and mixing processes would occur less effectively with the 45 deg swirler. Fuel and air are injected directly downward into the primary zone, reducing the time for the three processes. Fuel-air jets around the perimeter are directed toward the wall and could lead to fuel droplet impingement on the walls. The lack of recirculation could reduce the amount of shear that enhances fuel atomization and the degree of mixing that leads to uniform lean combustion.

60 Deg Clockwise and Counterclockwise Swirlers Case Study.. Comparisons are also made for 60 deg swirlers where the center swirler twists in the counterclockwise direction, while the outer eight remain clockwise. The results show a quite distinctive swirling profile compared to the two swirler arrangements

presented above. The counter-rotation causes much higher velocity in the combustor around the central circle, as the xy -plane plots show in Fig. 10. These two plots also show that the RSTM consistently predicts a more intense and vigorous velocity profile in comparison to the $k-\epsilon$ model. The reversed twisting of the central swirler also brings about stronger and deeper recirculations, which are evident when comparing Fig. 6 with Fig. 11 where reversed flow can still be observed below the $z=18$ mm plane.

An apparent outcome of this arrangement is that reversing the central swirler rotation enhances the strength of swirl flow, not only in the vicinity of the central circle, but also throughout the entire region inside the combustor. This can be explained by carefully studying the nature of the interactions among the swirling flows emanated from the nine swirlers. In the vicinity where the central circle meets its neighboring four, velocity is enhanced and the swirl gains strength, since the flows are in the same directions in these regions. Unlike the co-rotating swirler configuration dis-

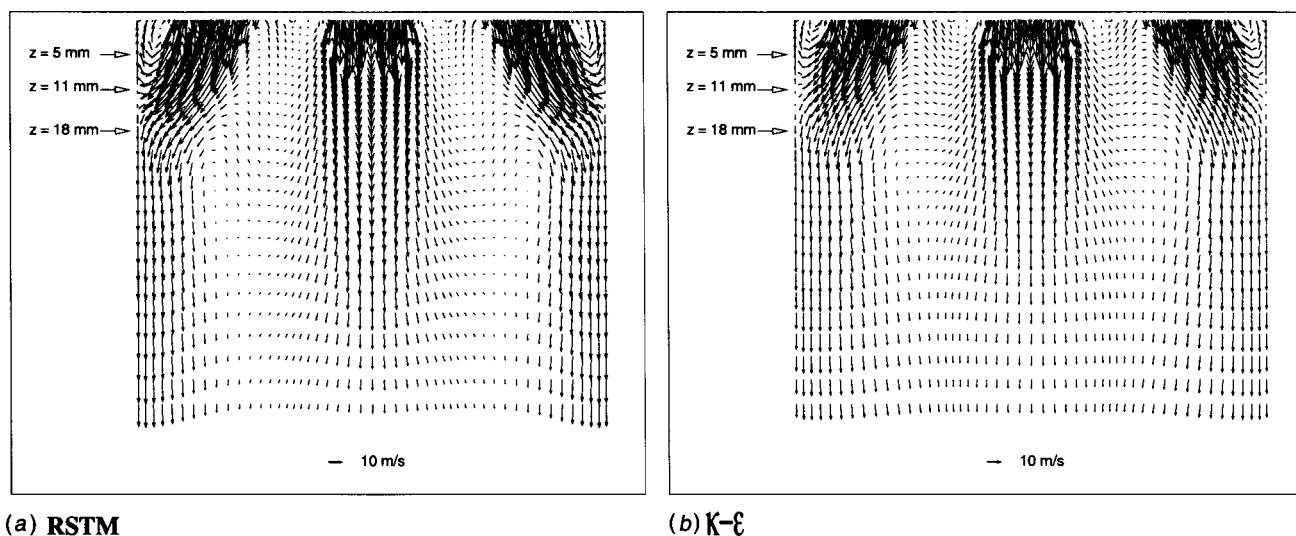


Fig. 9 RSTM velocity on y_0-y_0 plane for 45 deg swirler: (a) Reynolds stress turbulence model, (b) $k-\epsilon$

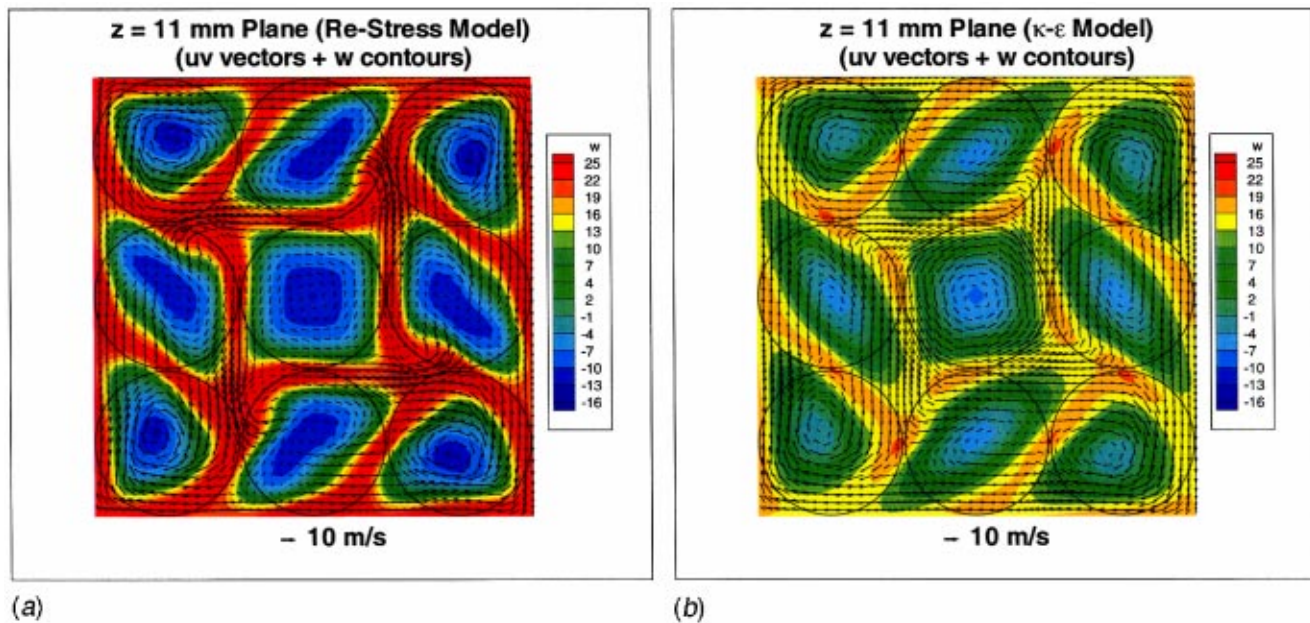


Fig. 10 Results for 60 deg clockwise and counterclockwise swirler at $z = 11$ mm plane: (a) Reynolds stress turbulence model, (b) $k-\epsilon$

cussed above where “shear canceling” occurs surrounding the central circle, this cw+ccw arrangement produces a more vigorous swirling flow under the same inlet boundary conditions.

It is also noted that, with this swirler arrangement, the recirculations exist at different locations from those of the co-rotating configuration. A comparison between the y_0 - y_0 cutting planes in Fig. 6 and Fig. 11 indicates that the reversed flow exhibits different patterns between the arrangements. Figure 6 shows four recirculation regions, all of which are located near the junctions of the swirlers and away from the vertical combustor walls. In Fig. 11, however, the four recirculation zones exist further away from one another. In particular, the two outer recirculations reside nearer the combustor walls. Such dissimilarity in the spatial locations of the recirculation regions may lead to different combustor designs, in terms of thermal protection on the combustor walls, flame stability, and pollutant emission control. These issues will be considered and addressed when flows with chemistry are studied.

Conclusion

The performance of the SSG Reynolds stress closure model using the KIVA-3V computer code is presented for chemically nonreacting flow simulations inside an LDI combustor. The RSTM performed better than the standard $k-\epsilon$ model when comparisons were made with experiment for 60 deg co-rotating swirlers. The primary improvements were the capturing of recirculation regions and the accompanying reversed flow. These flow characteristics were predicted to be stronger and more distinct with the RSTM. The results reaffirm a previous research statement that the RSTM can out-perform the $k-\epsilon$ model when swirling flows and recirculations are present. The results from the 45 deg swirler type indicated that the choice of swirler also plays an important role in the production of swirl and recirculation. The results from the 60 deg swirlers with reversed twisting angle at the center were also studied. This set of results clearly demonstrated the advantage the

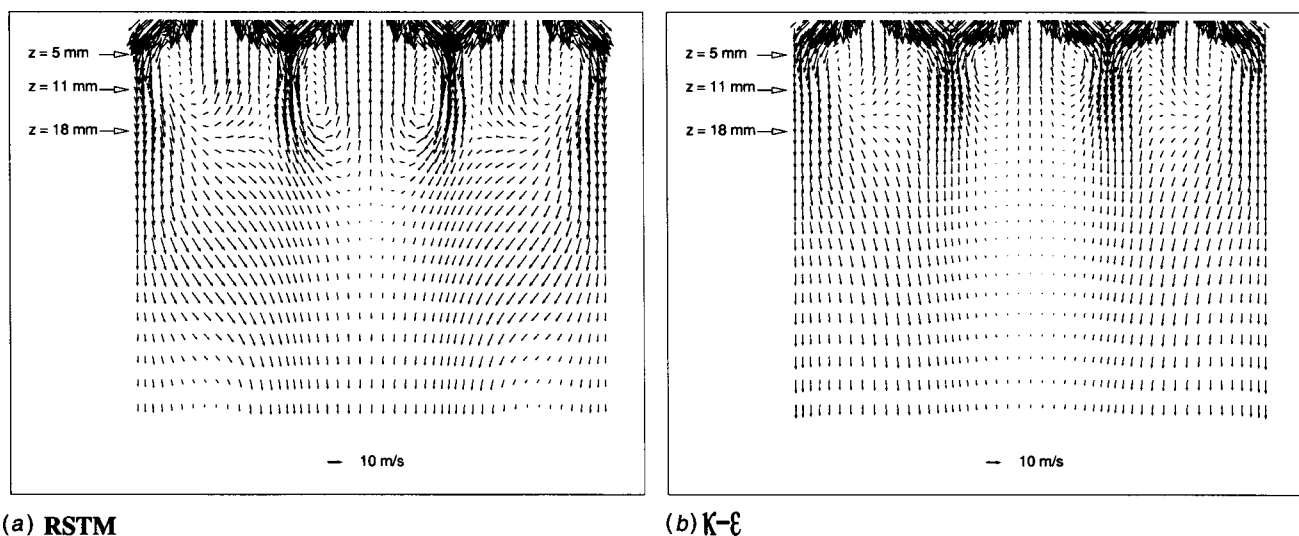


Fig. 11 RSTM velocity on y_0 - y_0 plane for 60 deg cw+ccw swirler: (a) Reynolds stress turbulence model, (b) $k-\epsilon$

RSTM has over the $k-\varepsilon$ model for parametric studies. Since in the RSTM each component of the Re-stresses is computed individually, no ad hoc modification is required when the problem geometry or flow condition is altered. In addition, the computation time for the RSTM increased only slightly in comparison to the $k-\varepsilon$ model. The $k-\varepsilon$ model required 0.5039 msec/cycle/grid point, whereas the RSTM required 0.5649 msec/cycle/grid point. In summary, RSTM has become a favorable method for providing predictions of turbulent swirling flow within a combustor. Success in these aspects should aid in providing successful predictions when chemically reactive flows are studied.

Quantitative comparisons showing the improvement experienced with the RSTM can be found in Refs. [25] and [26], which describe simulation results for an LDI combustor with discrete-jet swirlers using the RSTM. The comparisons were made to LDV data, [27], as well as results from the National Combustor Code using a cubic nonlinear $k-\varepsilon$ model, [28], tuned for swirl flow calculation.

Acknowledgments

The authors wish to thank NASA for their financial assistance under NASA Grant Contract #: NAG3-2088. Experimental information provided by D. L. Bulzan of NASA GRC is greatly appreciated.

Nomenclature

C_D	= discharge coefficient
d	= swirler inlet diameter
k	= turbulence kinetic energy
k_{in}	= inlet turbulence kinetic energy
p	= fluid pressure
$\tilde{u}_i \tilde{u}_j$	= Reynolds stress tensor
V_{in}	= inlet velocity, Eq. (13)
x_k	= coordinate vector
ε	= specific dissipation rate of turbulence kinetic energy
ε_{in}	= inlet dissipation rate of turbulence kinetic energy
ρ	= density

Subscripts

i, j, k = tensors

Superscripts

- = Reynolds averaged quantities
- ~ = Favre averaged quantities

References

- [1] Koff, B. L., 1994, "Aircraft Gas Turbine Emissions Challenge," ASME J. Eng. Gas Turbines Power, **116**, pp. 474–477.
- [2] Talpallikar, M. V., Smith, C. E., Lai, M. C., and Holdeman, J. D., 1992, "CFD Analysis of Jet Mixing in Low NO_x Flametube Combustors," ASME J. Eng. Gas Turbines Power, **114**, pp. 416–424.
- [3] Cline, M. C., Micklow, G. J., Yang, S. L., and Nguyen, H. L., 1995, "Numerical Analysis of the Flow-Fields in a Staged Gas Turbine Combustor," J. Propul. Power, **11**(5), pp. 894–898.
- [4] Yang, S. L., Chen, R., Cline, M. C., and Nguyen, H. L., 1992, "Numerical Simulation of a Low-Emission Gas Turbine Combustor Using KIVA-II," Int. J. Numer. Methods Fluids, **15**, pp. 865–881.
- [5] Micklow, G. J., Roychoudhury, S., Nguyen, H. L., and Cline, M. C., 1993, "Emissions Reduction by Varying the Swirler Air Flow Split in Advanced Gas Turbine Combustors," ASME J. Eng. Gas Turbines Power, **115**, pp. 563–569.
- [6] Kumakura, M. H., Sasaki, M., and Suzuki, D., 1993, "Low NO_x Combustor for Automotive Ceramic Gas Turbine-Reliability Assurance," ASME J. Eng. Gas Turbines Power, **115**, pp. 547–553.
- [7] Cowell, L. H., and Smith, K. O., 1993, "Development of a Liquid-Fueled, Lean-Premixed Gas Turbine Combustor," ASME J. Eng. Gas Turbines Power, **115**, pp. 554–562.
- [8] Leonard, G., and Stegmaier, J., 1994, "Development of an Aeroderivative Gas Turbine Dry Low Emissions Combustion System," ASME J. Eng. Gas Turbines Power, **116**, pp. 542–546.
- [9] Tacina, R. R., 1990, "Low NO_x Potential of Gas Turbine Engines," AIAA Paper No. 90-0550.
- [10] Gran, I. R., Ertesvag, I. S., Magnussen, B. F., 1997, "Influence of Turbulence Modeling of Predictions of Turbulent Combustion," AIAA J., **35**(1), pp. 106–110.
- [11] McQuirk, J. J., and Palma, J. M. L. M., 1993, "The Flow Inside a Model Gas Turbine Combustor: Calculations," ASME J. Eng. Gas Turbines Power, **115**, pp. 594–602.
- [12] Rizk, N. K., and Mongia, H. C., 1993, "Three-Dimensional Gas Turbine Combustor Emissions Modeling," ASME J. Eng. Gas Turbines Power, **115**, pp. 603–611.
- [13] Relation, H. L., Battaglioli, J. L., and Ng, W. F., 1998, "Numerical Simulations of Nonreacting Flows for Industrial Gas Turbine Combustor Geometries," ASME J. Eng. Gas Turbines Power, **120**, pp. 460–467.
- [14] Sloan, D. G., Smith, P. J., and Smooth, L. D., 1986, "Modeling of Swirl in Turbulent Flow Systems," Prog. Energy Combust. Sci., **12**, pp. 103–250.
- [15] Tsao, J. M., and Lin, C. A., 1999, "Reynolds Stress Modeling of Jet and Swirl Interaction Inside a Gas Turbine Combustor," Int. J. Numer. Methods Fluids, **29**(4), pp. 451–464.
- [16] Lin, C. A., and Lu, C. M., 1994, "Modeling Three-Dimensional Gas-Turbine Combustor Model Flow using a Second Moment Closure," AIAA J., **32**(7), pp. 1416–1422.
- [17] Gibson, M. M., and Launder, B. E., 1978, "Ground Effects on Pressure Fluctuations in the Atmospheric Boundary Layers," J. Fluid Mech., **86**, pp. 491–551.
- [18] Yang, S. L., Peschke, B. D., and Hanjalic, K., 2000, "Second-Moment Closure Model for IC Engine Flow Simulation Using KIVA Code," ASME J. Eng. Gas Turbines Power, **122**(2), pp. 355–365.
- [19] Speziale, C. G., Sarkar, S., and Gatski, T. B., 1991, "Modeling the Pressure-Strain Correlation of Turbulence: An Invariant Dynamical Systems Approach," J. Fluid Mech., **227**, pp. 245–272.
- [20] Amsden, A. A., 1997, "KIVA-3V: A Block-Structured KIVA Program for Engines With Vertical or Canted Valves," Los Alamos National Laboratory, LA-13313-MS, Los Alamos, NM.
- [21] Im, K.-S., Lai, M.-C., and Tacina, R., 1998, "A Parametric Spray Study of the Swirler/Venturi Injectors," AIAA Paper No. 98-3269.
- [22] Bachalo, W. D., and Houser, M. J., 1984, "Phase/Doppler Spray Analyzer for Simultaneous Measurements of Drop Size and Velocity Distributions," Opt. Eng., **23**(5), pp. 583–590.
- [23] Bulzan, D. L., 1995, "Structure of a Swirl-Stabilized Combusting Spray," J. Propul. Power, **11**(6), pp. 1093–1102.
- [24] Gupta, A. K., Lilley, D. G., and Syred, N., 1984, *Swirl Flows*, Abacus Press, Tunbridge Wells, England, Chap. 4.
- [25] Yang, S. L., Siow, Y. K., Teo, C. Y., Tacina, R. R., Iannetti, A. C., and Penko, P. F., 2003, "Numerical Study of LDI Combustor With Discrete-Jet Swirlers Using Reynolds Stress Model," ASME J. Eng. Gas Turbines Power, submitted for publication.
- [26] Teo, C. Y., Siow, Y. K., and Yang, S. L., 2001, "Flowfield Study of LDI Combustor With Discrete-Jet Swirler Using Re-Stress Model," AIAA Paper No. 2001-3424.
- [27] Jeng, S.-M., Cai, J., and Tacina, R. R., 2001, "Multi-Swirler Aerodynamics: Experimental Measurements," AIAA Paper No. 2001-3574.
- [28] Iannetti, A., Tacina, R. R., Cai, J., and Jeng, S.-M., 2001, "Multi-Swirler Aerodynamics: CFD Predictions," AIAA Paper No. 2001-3575.

Air Fuel Ratio Estimation Using In-Cylinder Pressure Frequency Analysis

N. Cavina

F. Ponti

DIEM—University of Bologna,
Viale Risorgimento, 2,
40136 Bologna, Italy

This paper presents an original approach to estimate the air-fuel ratio (AFR) of the mixture that burned inside a given cylinder of a spark-ignited (SI) internal combustion engine, using the information hidden in the corresponding in-cylinder pressure signal. In modern closed-loop fuel injection control strategies, the feedback signal is usually given by one (or more) heated exhaust gas oxygen (HEGO) sensor(s), mounted in the exhaust manifold(s). The information that such sensors give is related to the stoichiometry of the mixture that burned inside the cylinders. The HEGO sensor is not able to evaluate the AFR value precisely, being only able to determine whether the mixture was rich or lean. This information is sufficient to allow the implementation of a closed-loop strategy for injection time control. Generally speaking, such strategy could be improved in terms of readiness and precision by directly measuring (or by estimating) the actual AFR. Universal exhaust gas oxygen (UEGO) sensors are still considered expensive and their use is mostly limited to laboratory and racing applications, even if some automotive manufacturers have started installing such sensors on board passenger cars, as part of an effort to comply with ULEV (ultra low emission vehicles) regulations. For this reason the idea of estimating AFR values from other signals has received great attention in the past few years. A new approach based on in-cylinder pressure frequency analysis is presented here. [DOI: 10.1115/1.1563242]

Introduction

Precise air-fuel ratio (AFR) control, also during transient operating conditions, is the primary goal of modern spark-ignited (SI) engine control strategies, [1–6]. Being able to maintain a stoichiometric AFR allows high conversion efficiencies in the catalytic converter, thus permitting low polluting emissions production.

The possibility of sensing the AFR through a heated exhaust gas oxygen (HEGO) sensor plays a very important role for the definition of a closed-loop injection control strategy. Anyway, HEGO sensors can only determine if the mixture that was formed during injection and intake phases, and then burned inside the cylinders, was rich or lean, without quantifying AFR deviations from the stoichiometric value. Obviously, information about the real AFR value could improve control strategies performance (in terms of readiness, precision, and robustness), [2–8].

The AFR estimation approach developed here is based on in-cylinder pressure waveform frequency analysis: This allows individual cylinder AFR control by evaluating how much rich or lean the mixture was inside each cylinder. Furthermore, a second benefit on the control system bandwidth has to be considered: the use of in-cylinder pressure waveform allows a slight reduction in the delay between injection actuation and the instant at which the corresponding AFR value feedback signal is sent to the ECU. In fact, it has to be underlined that, in modern engine control strategies, the fuel injection usually ends before the intake valves open. Therefore, the decision about the amount of fuel to be injected has to be taken a few milliseconds before. After the injection ends and after the intake valves close, compression and expansion strokes occur; then the exhaust gases that burned inside the cylinder pass through the exhaust manifold where the HEGO (or universal exhaust gas oxygen (UEGO)) sensor is placed. All these phases require time, and therefore there exists a delay be-

tween the injection actuation and the sensing of the AFR related to that injection, [9]. This delay could be slightly reduced by estimating the AFR using in-cylinder pressure measurements, because this estimation could be performed at the end of each expansion stroke, thus avoiding the mass transport delay necessary for the exhaust gases to reach the HEGO (or UEGO) sensor.

In this work, in-cylinder pressure has been measured in an engine test cell together with AFR (using a UEGO sensor), in order to identify and validate the estimation algorithm. The application of this approach on board a vehicle is obviously possible only if in-cylinder pressure is measured, thus making the strategy hardly implementable, at this time, for costs and durability problems related to modern in-cylinder pressure sensors. Nevertheless, previous studies demonstrated the possibility of reconstructing the low frequency content of in-cylinder pressure signal from engine speed fluctuations, [10–13], and other studies are being done in order to develop a new kind of low-cost pressure sensors that could be installed on board the vehicle, [14–16]. The use of the algorithm on board the vehicle will be possible if further studies will demonstrate that its application is feasible also by using the pressure waveform frequency content reconstructed from engine speed fluctuations, or if new low-cost pressure sensors will be placed into market.

In the following, the theoretical fundamentals that allow understanding of where AFR-related information is mostly evident in an in-cylinder pressure signal are presented. The experimental tests that have been conducted to verify, identify, and validate this approach are also briefly discussed.

Theoretical Fundamentals

The theoretical fundamentals upon which the approach is based are presented in this section. These theoretical fundamentals are used to understand if and how AFR variations can influence in-cylinder pressure waveforms; this is done by working in the angular frequency domain, and thus analyzing the in-cylinder pressure harmonic components.

In order to understand how in-cylinder pressure harmonic com-

Contributed by the Internal Combustion Engine Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the ICE Division, Feb. 2001; final revision received by the ASME Headquarters, May 2002. Associate Editor: D. Assanis.

ponents are influenced by AFR variations, the correlation between in-cylinder pressure and indicated torque mean value over an engine cycle is analyzed.

The indicated torque for a Z-cylinder engine can be expressed, in the angular domain, as

$$T_i(\theta) = A_p r \sum_{m=1}^Z p_m(\theta) f_m(\theta). \quad (1)$$

Considering for simplicity the contribution of the m th cylinder and focusing the attention over an engine cycle, during which N samples have been acquired, Eq. (1) becomes

$$T_{im}(n) = A_p r p_m(n) f_m(n) \quad n = 1, \dots, N \quad (2)$$

where n stands for the n th sample acquired in correspondence of the angular position $n \cdot \Delta\theta$.

Using Eq. (2), indicated torque evaluation for the m th cylinder is possible if the corresponding in-cylinder pressure is known. In order to evaluate the indicated torque mean value over an engine cycle, only few information relative to in-cylinder pressure are needed. This is due to the fact that the crank-slider function is the sum of two periodic terms, the first one with period 2π , the second one π (see the Nomenclature). A consequence of this is, for example, that a constant in-cylinder pressure over a whole engine cycle results in an indicated torque on the crankshaft having mean value equal to zero. This yields to the conclusion that the in-cylinder pressure cycle mean value has no influence on indicated torque cycle mean value evaluation. Further investigations on the relationship existing between in-cylinder pressure and indicated torque demonstrate that other in-cylinder pressure harmonic components do not have any influence on indicated torque cycle mean value, as it will be shown in the following, [10].

The discrete Fourier transform (DFT) of Eq. (2) yields

$$\begin{aligned} T_{im}(j\Omega_k) &= A_p r P_m(j\Omega_k) * F_m(j\Omega_k) \\ &= A_p r \sum_{h=-N/2}^{N/2} P_m(j\Omega_h) F_m(j\Omega_{k-h}) \\ k &= -N/2, \dots, 0, \dots, N/2, \quad \Omega_k = 2\pi k/N \end{aligned} \quad (3)$$

where the symbol $*$ indicates the convolution between the DFTs of the in-cylinder pressure $P_m(j\Omega_k)$ and of the crank-slider mechanism function $F_m(j\Omega_k)$.

Since $P_m(j\Omega_k)$ and $F_m(j\Omega_k)$ are distinct only over a range of N consecutive values of k , the summation of Eq. (3) has been limited to the terms in the range $-N/2$ to $N/2$.

The indicated torque mean value can be obtained from Eq. (3) for $k=0$:

$$\bar{T}_{im} = T_{im}(j0) = A_p r \sum_{h=-N/2}^{N/2} P_m(j\Omega_h) F_m(j\Omega_{-h}) \quad (4)$$

The spectral analysis of the function $f_m(n)$ shows that only two angular frequency components (the second and the fourth over an engine cycle) have an amplitude significantly different from zero, while all the other components show zero or very close to zero amplitude values, as can be seen in Table 1.

This observation allows neglecting all the terms of the summation in Eq. (4) that present an amplitude close to zero. Noting that $F_m(j\Omega_{-h})$ is the complex conjugate of $F_m(j\Omega_h)$, it is finally possible to obtain

$$\begin{aligned} \bar{T}_{im} &= 2A_p r [\operatorname{Re}\{P_m(j\Omega_4)F_m(j\Omega_4)^*\} + \operatorname{Re}\{P_m(j\Omega_2)F_m(j\Omega_2)^*\}] \\ &= 2A_p r \sum_{q=1}^2 [\operatorname{Re}\{P_m(j\Omega_{2q})F_m(j\Omega_{2q})^*\}]. \end{aligned} \quad (5)$$

It is evident from Eq. (5) that, for the evaluation of the indicated torque mean value over an engine cycle, it is sufficient to

Table 1 Amplitude of the angular frequency components of the crank slider mechanism function $f_m(\theta)$ for typical values of λ .

Angular Frequency Component Order	Frequency Component Amplitude Value		
	$\lambda=0.2$	$\lambda=0.3$	$\lambda=0.4$
0	0	0	0
± 1	0	0	0
± 2	0.50	0.50	0.50
± 3	0	0	0
± 4	0.05	0.08	0.10
± 5	0	0	0
± 6	0	0	0
± 7	0	0	0
± 8	negligible $0.26 \cdot 10^{-3}$	negligible $0.90 \cdot 10^{-3}$	negligible $2.27 \cdot 10^{-3}$
$\pm > 8$	0	0	0

know the second and fourth in-cylinder pressure angular frequency components over the engine cycle, for the cylinder taken into account.

In the following, the in-cylinder pressure waveform over an engine cycle is considered as being split into two components: the first one corresponding to the compression and expansion strokes, $p_c(n)$, the second one corresponding to the exhaust and intake strokes, $p_p(n)$. The subscript m (referring to the m th cylinder) is omitted in the following for ease of understanding.

In order to evaluate the indicated torque mean value over each of the two engine cycle portions considered, Eqs. (2), (3), and (4) can still be considered, the only difference being that in this case the number of samples in the angular range taken into account is $N/2$ and therefore $n = 1, \dots, N/2$ and $k = -N/4, \dots, 0, \dots, N/4$. Equation (4) could then be rewritten, for example, for the evaluation of the indicated torque mean value during the compression and expansion strokes T_{ic} .

$$\bar{T}_{ic} = T_{ic}(j0) = A_p r \sum_{h=-N/4}^{N/4} P_c(j\Omega_{k-h}) F_c(j\Omega_{-h}) \quad (6)$$

It has to be noted that the relationship existing between $\bar{T}_{ic}$, T_{ip} and T_{im} can be expressed as $T_{im} = (T_{ic} + T_{ip})/2$, since the angular interval over which T_{im} is evaluated is twice the one over which T_{ic} and T_{ip} are evaluated.

In the following, the attention has been focused on the half engine cycle during which compression and expansion strokes occur (and thus the analysis has been focused on p_c), because AFR effects have shown to be particularly evident during this half-cycle.

The crank-slider mechanism function spectral analysis shows that, over half a cycle, the frequency components significantly different from zero are, in this case, the first and the second. Equation (6) can then be rewritten as

$$\bar{T}_{ic} = 2A_p r \sum_{q=1}^2 [\operatorname{Re}\{P_c(j\Omega_q)F_c(j\Omega_q)^*\}]. \quad (7)$$

Over the half-cycle taken into account $\operatorname{Re}\{F_c(j\Omega_1)\} = \operatorname{Re}\{F_c(j\Omega_2)\} = 0$, since $f_m(\theta)$ is purely sinusoidal. Equation (7) can therefore be simplified as

$$\bar{T}_{ic} = 2A_p r \sum_{q=1}^2 [\operatorname{Im}\{P_c(j\Omega_q)\} \cdot \operatorname{Im}\{F_c(j\Omega_q)\}]. \quad (8)$$

As shown in Eq. (8), only the first two imaginary parts of the in-cylinder pressure harmonic components play a role for $\bar{T}_{ic}$ evaluation. In order to understand how p_c frequency components can be influenced by AFR variations, a specific set of experimental tests has been designed.

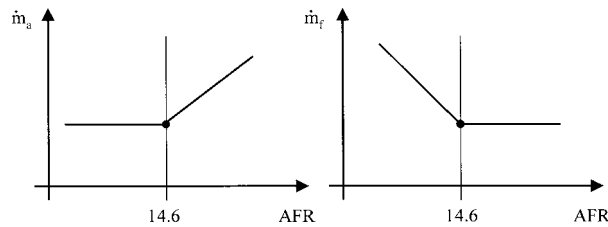


Fig. 1 Controlled air and fuel mass-flow rates variations to vary AFR values while maintaining T_{ic} constant

It is worth spending some words to explain how these tests have been performed. Spark advance (SA) has been controlled at a constant value; engine speed has been maintained constant by the engine test-cell controller. At this point, throttle opening and fuel injection time have been chosen in order to obtain tests with a constant T_{ic} value and different AFR values.

Since T_{ic} remains constant, also the fuel mass burned inside the cylinders can be considered, in first approximation, constant. The amount of fuel burned is approximately equal to the amount of fuel injected if the AFR is lean, while it is related to the amount of air introduced into the cylinder if the AFR is rich. With respect to stoichiometric conditions, then, it is possible to enrich the AFR without changing the amount of fuel burned, by increasing the amount of fuel injected. On the contrary, imposing a lean AFR while maintaining the same amount of fuel burned is possible by injecting the same amount of fuel while increasing the mass of air entering the cylinder. This means that it is possible to obtain a lean AFR without changing the amount of fuel burned by slightly opening the throttle, while maintaining the same injection time as the corresponding stoichiometric operating condition.

Figure 1 summarizes air and fuel mass-flow rates variations laws for these tests taking as a reference the air and fuel mass flow rates obtained for stoichiometric conditions.

Air and fuel mass-flow rates variations suggested in Fig. 1 should allow maintaining constant the amount of fuel burned inside the cylinders. This allows maintaining the value of T_{ic} approximately constant but does not guarantee that T_{ic} is exactly constant. In fact, flame speed propagation inside the cylinder is not the same for all the AFR values, slowing down for leaner AFRs. This decreases the value of T_{ic} obtained for leaner AFRs if the air and fuel mass-flow rates variations are those reported in Fig. 1. This means that Fig. 1 can only play an orientation role in performing a set of tests having constant T_{ic} , being necessary a further adjustment when performing the experiments in the engine test cell.

For each SA, engine speed and T_{ic} operating condition, it is then possible to obtain a group of tests characterized by different AFR values by changing fuel injection time and throttle opening, in a way similar to what suggested in Fig. 1. In the following, one of these groups of tests (obtained by imposing constant SA, engine speed and T_{ic}) will be considered as an example. The same considerations are of course valid for every group of tests of the same type.

AFR variations do influence the pressure waveforms relative to the tests taken into account, and therefore the corresponding angular frequency harmonic components assume different values. However, all the operating conditions taken into account in this group of tests present the same T_{ic} .

This means that as the AFR changes, the first two in-cylinder pressure frequency components imaginary parts are bound to vary maintaining $2A_p r \sum_{q=1}^2 [\text{Im}\{P_c(j\Omega_q)\} \cdot \text{Im}\{F_c(j\Omega_q)\}]$ constant, since T_{ic} is maintained constant. For this reason, the effects of AFR variations on in-cylinder pressure waveform will be particu-

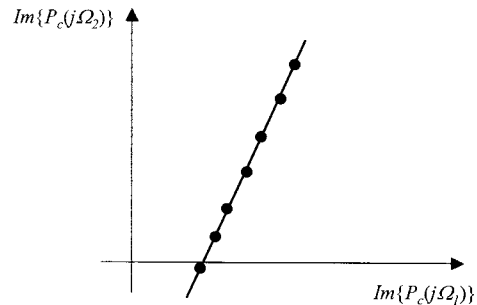


Fig. 2 P_c harmonic components imaginary parts, for a group of operating conditions at the same engine speed, T_{ic} and SA

larly evident by looking at the real parts of the first two in-cylinder pressure harmonic components, called in the following F_1 and F_2 .

This observation does not mean that no information about AFR variation is contained in the imaginary parts of the first two in-cylinder pressure harmonic components. In fact, the constant value of T_{ic} does not imply a constant value for the two imaginary parts.

In a plane where the two imaginary parts are the coordinates, Eq. (8) describes in fact a straight line upon which the points representing the tests lie (see Fig. 2), since $\text{Im}\{F_c(j\Omega_1)\}$ and $\text{Im}\{F_c(j\Omega_2)\}$ are, of course, constant for a given engine.

The harmonic components imaginary parts are free to move along the straight line of Fig. 2, without affecting the value of T_{ic} . The motion along the straight line will be obviously related to AFR changes; in order to obtain another parameter to be used for AFR evaluation, the amplitude of this motion can be evaluated. To do this, an axes rotation has to be applied to the axes of Fig. 2, in order to make the direction of one of the two new axes the same as the straight line one (see Fig. 3).

This rotation allows the identification of a third parameter, F_3 , to be used for AFR estimation, the first two parameters being the real parts of the first two in-cylinder pressure harmonic components. Using simple geometric considerations it is possible to obtain

$$\begin{cases} F_3 = \text{Im}\{P_c(j\Omega_1)\} \cdot \text{Im}\{F_c(j\Omega_2)\} - \text{Im}\{P_c(j\Omega_2)\} \cdot \text{Im}\{F_c(j\Omega_1)\} \\ F_4 = \text{Im}\{P_c(j\Omega_1)\} \cdot \text{Im}\{F_c(j\Omega_1)\} + \text{Im}\{P_c(j\Omega_2)\} \cdot \text{Im}\{F_c(j\Omega_2)\} \end{cases} \quad (9)$$

This theoretical approach allows then to identify three parameters, F_1 , F_2 , and F_3 , derived by in-cylinder pressure frequency analysis, and containing information on AFR variations that can be used for its estimation.

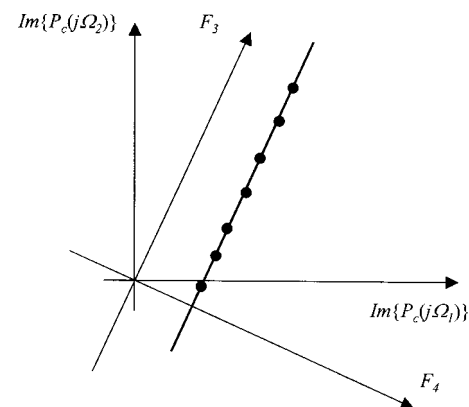


Fig. 3 Axes rotation

Table 2 Main engine and sensors characteristics

Engine Characteristics	
Type	FIAT, multipoint inj., 4L
Displacement	2.0 liters
Cylinder bore	84 mm
Stroke	90 mm
Connecting rod length	145 mm
Maximum power	109 kW at 6200 rpm
Maximum torque	180 Nm at 4500 rpm
Sensors Characteristics	
UEGO sensor	HORIBA-MEXA 700 LAMBDA
In-cylinder pressure sensor	KISTLER 6051A

The fourth parameter F_4 is related to $\bar{T}_{ic}$ variations and can remain constant even if AFR changes: For this reason it has not been used for AFR estimation.

Experimental Tests

Some experimental tests have been conducted to identify and validate the AFR estimation methodology previously presented. The tests have been conducted running an SI engine in a test cell.

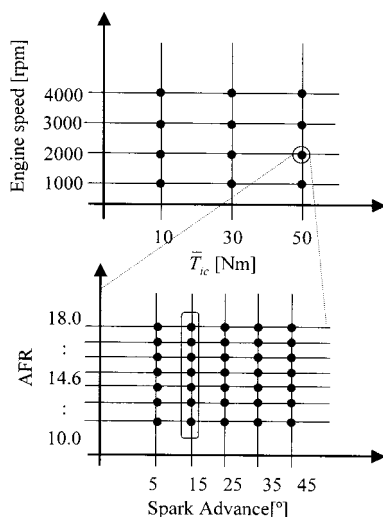
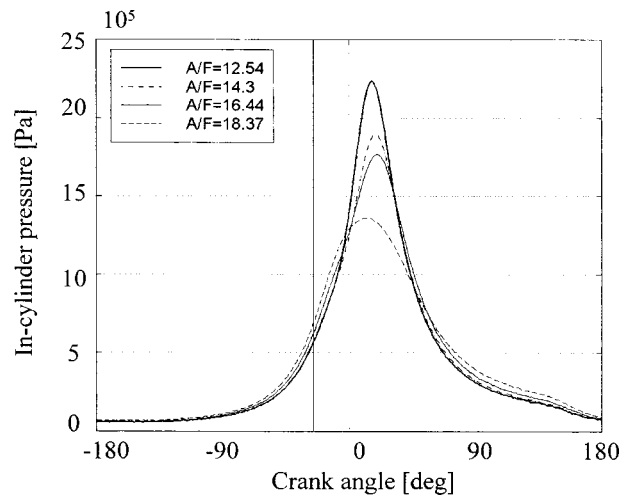
The main characteristics of the engine used and of the sensors employed are reported in Table 2.

The experimental tests have been conducted running the engine for nine different steady-state load-speed conditions. Each operating condition has been further investigated imposing different SA and AFR values.

Figure 4 summarizes the way the tests have been run.

For each steady-state engine operating condition, in-cylinder pressure and AFR have been acquired for 100 consecutive engine cycles. A preliminary analysis has then been performed evaluating the mean in-cylinder pressure waveform over an engine cycle and the mean AFR value, thus obtaining a single pressure waveform and a single AFR value for each operating condition considered.

As already mentioned in the previous section, in order to understand how different AFRs can influence the pressure waveforms, tests with the same $\bar{T}_{ic}$ value, engine speed and SA have been considered. For this reason, the tests presented in Fig. 4 have been divided into groups showing the same $\bar{T}_{ic}$, engine speed and SA values. For example, the comparison between the pressure waveforms in cylinder 1 for tests at 30 Nm $\bar{T}_{ic}$ (it corresponds to

**Fig. 4 Investigated engine operating conditions****Fig. 5 Cylinder 1 pressure waveforms at $\bar{T}_{ic}$ 30 Nm, 2000 rpm, 25 deg SA and different AFR values**

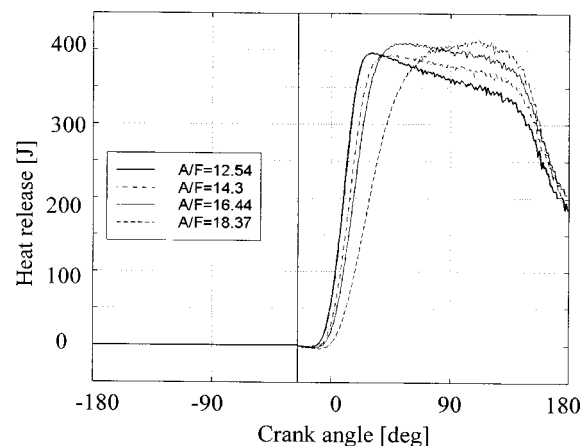
approximately 60 Nm $\bar{T}_i$ considering the contribution of all the cylinders), 2000 rpm and 25 deg SA is reported in Fig. 5, for the 360 deg centered around cylinder 1 combustion top dead center (TDC). The considerations that follow could be applied to other groups of tests characterized by constant SA, engine speed and $\bar{T}_{ic}$ values.

The vertical solid line in Fig. 5 indicates the angular position where the SA has been actuated for the tests considered. As it can be noted, the pressure waveforms show great differences for the crank angles that follow spark ignition; these differences seem to be mainly due to the flame propagation velocity values, decreasing with increasing AFR, as it is shown in Fig. 6, where net heat release has been determined from in-cylinder pressure measurements, according to [17,18]

$$\frac{dQ_n}{d\theta} = \frac{\gamma}{\gamma-1} p_m(\theta) \frac{dV(\theta)}{d\theta} + \frac{1}{\gamma-1} V(\theta) \frac{dp_m(\theta)}{d\theta}. \quad (10)$$

It can be noted that the smaller the AFR the higher is the flame propagation velocity, and also the pressure peak near TDC. At the same time the combustion becomes slower and the pressure waveform smoother as the AFR is decreasing.

The pressure waveforms show different trends also before spark ignition and start of combustion occur. These differences are related to the way the tests considered in Figs. 5 and 6 have been

**Fig. 6 Net heat release waveforms for tests at $\bar{T}_{ic}$ 30 Nm, 2000 rpm, 25 deg SA and different AFR values**

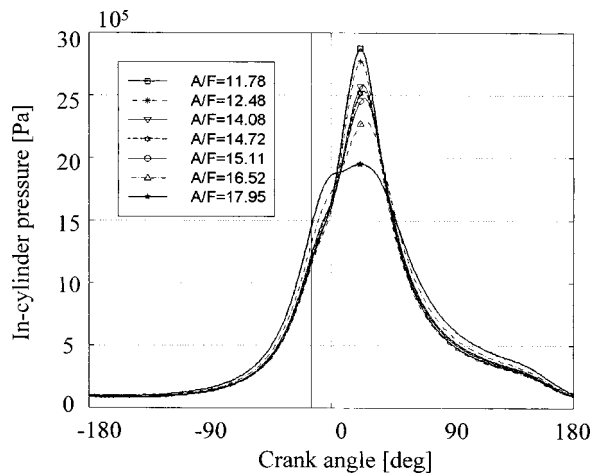


Fig. 7 Pressure waveforms at $\bar{T}_{ic}$ 50 Nm (corresponding to around $\bar{T}_i$ 100 Nm for all four cylinders), 2000 rpm, 15 deg SA, and different AFR values

obtained. As already mentioned in a preceding section, for the same group of tests it is possible to obtain a lean AFR without affecting the indicated torque value during compression and expansion strokes. This can be done maintaining the same injection time as the corresponding stoichiometric operating condition, while slightly opening the throttle. For this reason the tests during which a lean AFR has been imposed, show a manifold pressure higher than the corresponding tests with stoichiometric and rich AFRs, and therefore a higher in-cylinder pressure during intake and compression strokes.

Estimation Procedure Identification and Validation

The theoretical approach developed in previous sections allowed identifying three parameters, F_1 , F_2 , and F_3 , derived by in-cylinder pressure frequency analysis, that contain information on AFR variations and that can be used for its estimation. This section shows how these parameters have been used to develop an AFR estimation algorithm.

To do this, the group of tests shown in Fig. 7 and encircled in Fig. 4 has been considered here as an example. The same considerations can be applied to any other group of tests obtained by imposing constant values of engine speed, SA and T_{ic} .

Air and fuel mass-flow rate variations for the tests of Fig. 7 are reported in Fig. 8. As it can be seen these variations are similar, as expected, to those reported in Fig. 1.

Reporting on a diagram the imaginary parts of the first two in-cylinder pressure frequency components, the points representing the tests taken into account lie, as expected, on a straight line (see Fig. 9), identified by Eq. (8).

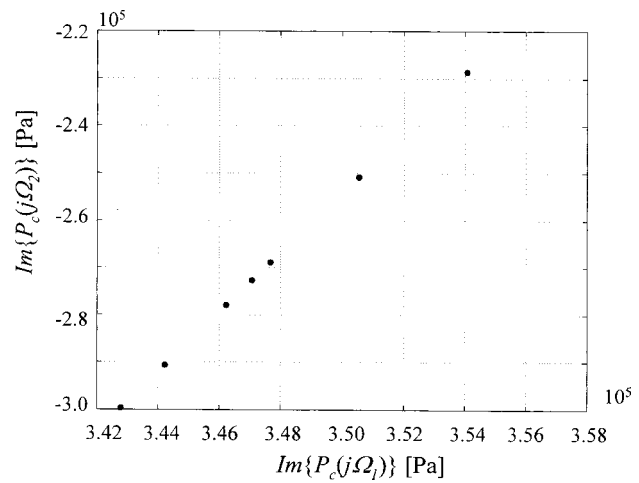


Fig. 9 Imaginary parts of the first two in-cylinder pressure frequency components

All the considerations made in the theoretical fundamentals section can now be applied to this group of tests, and then the parameters F_1 , F_2 , and F_3 can be used to extract information related to AFR variations.

In order to understand the nature of the relationship between AFR values and these parameters, the values of F_1 , F_2 , and F_3 , for the tests of Fig. 7 have been reported in Figs. 10, 11, and 12, as a function of AFR.

Figures 10, 11, and 12 show that the relationship between each F_i ($i=1,2,3$) parameter and AFR values can be reasonably considered linear. It is then possible to consider, for the group of tests taken into account, a relationship between AFR and each parameter of the type

$$AFR_{e1} = C_{01} + C_{11}F_1$$

$$AFR_{e2} = C_{02} + C_{12}F_2 \quad (11)$$

$$AFR_{e3} = C_{03} + C_{13}F_3.$$

Three AFR estimation procedures can therefore initially be considered, each of them using one of the parameters previously defined for AFR estimation. In order to identify the three estimation procedures over the whole engine operating range, all the experimental tests presented in the preceding sections (see Fig. 4) have been divided into groups characterized by the same engine speed, SA, and T_{ic} values. Each of such group of tests showed a linear relationship between AFR and the parameters F_1 , F_2 , and F_3 , of the type described in Eq. (11), even if each one is characterized by different values of the coefficients C_{0j} , C_{1j} .

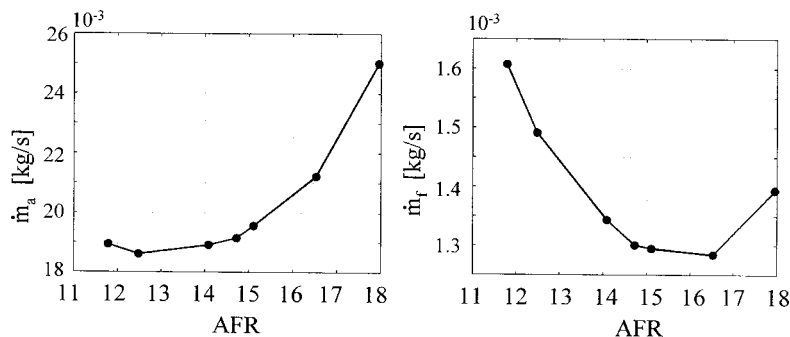


Fig. 8 Air and fuel mass-flow rates for the tests of Fig. 7

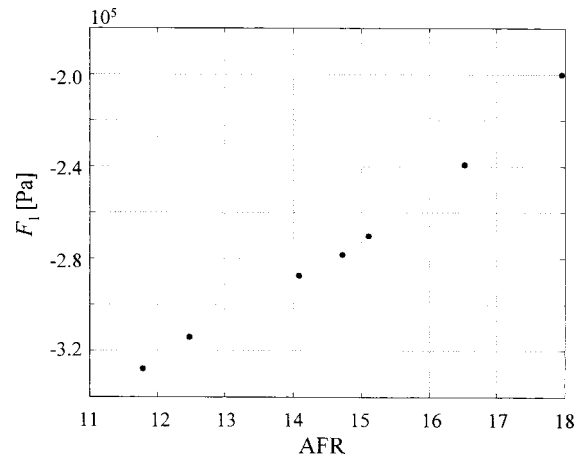


Fig. 10 F_1 variation for the tests of Fig. 7

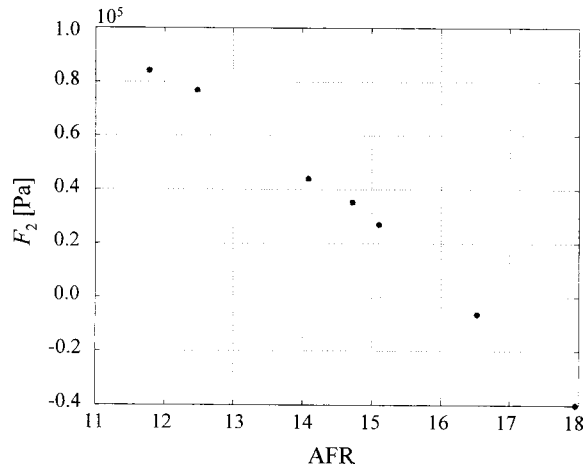


Fig. 11 F_2 variation for the tests of Fig. 7

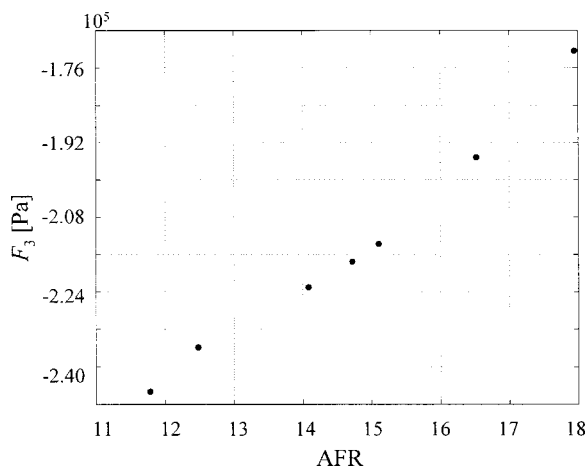


Fig. 12 F_3 variation for the tests of Fig. 7

In order to identify the coefficients C_{0j} , C_{1j} for each group of tests taken into account, a minimization process has been applied taking $\sigma^2 = \frac{\sum (AFR_e - AFR)^2}{\# \text{ of tests}}$ as the quantity to be minimized.

This allowed identifying the values of the coefficients C_{0j} , C_{1j} for each group of tests, and therefore for each speed, SA, T_{ic} engine operating condition considered. The coefficients

Table 3 F_1 , F_2 , and F_3 procedure AFR estimation mean absolute error

	AFR Estimation Mean Absolute Error
F_1	0.27
F_2	0.43
F_3	0.29
$F_1 \& F_2 \& F_3$	0.16

values have then been mapped as function of engine speed, SA, and T_{ic} over the whole engine operating range, thus completing the three identification processes.

$$AFR_{e1} = C_{01}(\omega, SA, \bar{T}_{ic}) + C_{11}(\omega, SA, \bar{T}_{ic})F_1$$

$$AFR_{e2} = C_{02}(\omega, SA, \bar{T}_{ic}) + C_{12}(\omega, SA, \bar{T}_{ic})F_2 \quad (12)$$

$$AFR_{e3} = C_{03}(\omega, SA, \bar{T}_{ic}) + C_{13}(\omega, SA, \bar{T}_{ic})F_3$$

This procedure validation has been conducted using an independent set of data, and comparing the procedure estimation with the AFR measured data. For each procedure the mean absolute AFR error has been evaluated and reported in Table 3.

F_1 and F_3 showed a lower value of the mean absolute error. This result is not uniform over the whole engine operating range: There exist, in fact, portions of the engine operating range where the absolute error obtained using F_3 is the lowest, and other portions where the lowest absolute error is that obtained using F_1 or F_2 .

In order to further improve the estimation methodology performance, the estimated AFR has also been expressed as a linear combination of the three F_i parameters. An equation of the type

$$AFR_e = C_0(\omega, SA, \bar{T}_{ic}) + C_1(\omega, SA, \bar{T}_{ic})F_1 + C_2(\omega, SA, \bar{T}_{ic})F_2 + C_3(\omega, SA, \bar{T}_{ic})F_3 \quad (13)$$

has been considered, where the coefficients C_0 , C_1 , C_2 , and C_3 are still mapped over the whole engine operating range as functions of T_{ic} , SA and engine speed, and identified by means of a minimization process.

This allowed reducing the mean AFR estimation error to 0.16 (see Table 3). A comparison between the measured AFR values and the corresponding estimated ones using the complete algorithm of Eq. (13) is finally reported in Fig. 13.

The results obtained show the capability of the procedure to estimate AFR values. The mean absolute AFR estimation error (around 1%) may still be too high to completely substitute the use of an oxygen sensor, even if its application together with a HEGO sensor should allow improving closed loop engine control performance, thus avoiding the use of a more expensive UEGO sensor.

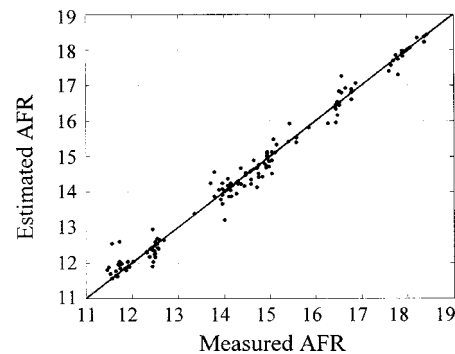


Fig. 13 AFR estimations versus corresponding AFR measurements

The procedure developed in this work has been identified and validated using pressure traces averaged over 100 cycles. A higher mean absolute error is then expected when performing AFR estimation using the cycle-to-cycle in-cylinder pressure traces. It should nonetheless be noted that the cyclic variability related to this signal slightly affects the lower order harmonics, its influence being more evident on the spectrum high frequency part. Therefore only a slight increase in the estimation error is expected, being the very lower part of the in-cylinder pressure spectrum used by the proposed AFR estimation algorithm.

The algorithm is also expected to be robust with respect to engine aging, but sensitive to the in-cylinder pressure sensor aging. The relationship between the AFR and the measured in-cylinder pressure is not expected to change with engine aging, while sensor degradation should result in different in-cylinder pressure traces for the very same engine operating conditions, thus affecting the estimation procedure precision.

In order to compare the AFR estimation algorithm presented here with those presented before in the literature, the algorithms proposed [2,4,5] have been identified and validated using the same data sets employed in this work. The best AFR mean absolute estimation error of 0.54 has been obtained using the algorithm proposed in [2], showing therefore a worse estimation accuracy compared with that obtained using the algorithm developed in this work.

Some words have to be finally spent about the computational time required to run the procedure. FFT algorithms usually require large amount of memory and are time-consuming. In this case it has to be underlined that the designed procedure does not require the evaluation of the whole in-cylinder pressure spectrum, but only of the first two angular frequency components over the compression and expansion phases, thus meaning that a low sampling rate is sufficient, and that specific advanced time-saving algorithms can be employed for the evaluation of such frequency components, [10].

Conclusions

The AFR estimation procedure has been defined by analyzing the relationship existing between AFR variations and the corresponding in-cylinder pressure waveforms. A theoretical approach has been introduced to identify the most relevant effects of AFR changes on the in-cylinder pressure frequency content, thus allowing the definition of an AFR estimation procedure.

The implementation on board a vehicle requires in-cylinder pressure knowledge. It will be possible if new low-cost in-cylinder pressure sensors will be placed into the market or if the low-frequency content of in-cylinder pressure can be precisely estimated. Previous works done by this group of research [10,13] demonstrated the possibility of estimating in-cylinder pressure using engine speed fluctuations. The precision obtained using that approach seems to be compatible with that required for the subsequent AFR estimation. Further investigations are being done in order to demonstrate the feasibility of the combination of these two approaches. This combination should allow estimating in-cylinder pressure from engine speed fluctuations and AFR values from the estimated pressure, thus allowing AFR estimation without the need for in-cylinder pressure sensors on board the vehicle.

Acknowledgments

The authors would like to thank Magneti Marelli for the support and, in particular, Dr. Gabriele Serra for his contribution in the development of the present work.

Nomenclature

- A_p = piston surface (m²)
- r = crank-slider mechanism radius (m)
- l = connecting rod length (m)
- θ = crank angle (rad)

- $\Delta\theta$ = crank angle sampling period (rad)
- $p_m(\theta)$ = relative pressure inside m th cylinder (Pa)
- $p_c(\theta)$ = relative pressure inside m th cylinder during compression and expansion strokes (Pa)
- $p_p(\theta)$ = relative pressure inside m th cylinder during intake and exhaust strokes (Pa)
- λ = r/l
- $f_m(\theta)$ = $\frac{\sin(\theta + \pi m) + \lambda \sin(2(\theta + \pi m))}{2\sqrt{1 - \lambda^2 \sin^2(\theta + \pi m)}}$ crank slider mechanism function for m th cylinder
- $T_i(\theta)$ = indicated torque of the engine (Nm)
- $T_{im}(\theta)$ = indicated torque for m th cylinder (Nm)
- $T_{ic}(\theta)$ = indicated torque for m th cylinder during compression and expansion strokes (Nm)
- $T_{ip}(\theta)$ = indicated torque for m th cylinder during intake and exhaust strokes (pumping torque) (Nm)
- Z = number of cylinders
- N = number of samples acquired in an engine cycle
- ω = engine angular velocity (rad/s)
- $\dot{m}_a$ = mean air mass-flow rate entering the cylinders (kg/s)
- $\dot{m}_f$ = mean fuel mass-flow rate entering the cylinders (kg/s)
- $\Omega_k = \frac{2\pi k}{N}$ = angular sample frequency (rad)
- k = angular sample frequency order
- F_1 = real part of the first in-cylinder pressure frequency component (Pa)
- F_2 = real part of the second in-cylinder pressure frequency component (Pa)
- F_3 = linear combination of the imaginary parts of the first two in-cylinder pressure frequency components (Pa)
- Q_n = net heat release (J)
- γ = specific heat ratio
- $V(\theta)$ = volume trapped inside the cylinder (m³)

References

- [1] Hendricks, E., and Sorenson, S. C., 1991, "SI Engine Controls and Mean Value Engine Modeling," SAE Technical Paper No. 910258.
- [2] Gassenfeit, E. H., and Powell J. D., 1989, "Algorithms for Air-Fuel Ratio Estimation Using Internal Combustion Engine Cylinder Pressure," SAE Technical Paper No. 890300.
- [3] Chang, C. F., Fekete, N. P., and Powell, J. D., 1993, "Engine Air-Fuel Ratio Control Using an Event-Based Observer," SAE Technical Paper No. 930766.
- [4] Houpt, P. K., and Andreadakis, S. K., 1983, "Estimation of Fuel/Air Ratio From Cylinder Pressure in Spark Ignited Engines," SAE Technical Paper No. 830418.
- [5] Fiorini, R. A., Gassenfeit, E. H., and Powell, J. D., 1988, "A/F Estimation From Internal Combustion Engine Cylinder Pressure Using Fast Fourier Transform Descriptors," *Proceedings of the American Control Conference*, **88**, Pt. 1-3, June 15-17, Atlanta, GA, American Automatic Control Council, Green Valley, AZ.
- [6] Powell, J. D., 1993, "Engine Control Using Cylinder Pressure: Past, Present and Future," *ASME J. Dyn. Syst., Meas., Control*, **115**, pp. 343-350.
- [7] Iwata, Y., Abe, S., Nasu, M., and Aoki, K., 1996, "Model-Based Air-Fuel Ratio Control With Exhaust Air-Fuel Ratio Sensor and Heated Exhaust Oxygen Sensor," *Fisita XXVI Congress*, Paper No. P16.50.
- [8] Kishi, N., Kikuchi, S., Seki, Y., Kato, A., and Fujimori, K., 1998, "Development of the High Performance L4 Engine ULEV System," SAE Technical Paper No. 980415.
- [9] Shafai, E., Roduner, C. A., and Geering, H. P., 1997, "On-line Identification of Time Delay in the Fuel Path of an SI Engine," SAE Technical Paper No. 970613.
- [10] Cavina, N., Ponti, F., and Rizzoni, G., 1999, "Fast Algorithm for On-Board Torque Estimation," SAE Technical Paper No. 1999-01-0541.
- [11] Azzoni, P., Moro, D., Ponti, F., and Rizzoni, G., 1998, "Engine and Load Torque Estimation With Application to Electronic Throttle Control," SAE Technical Paper No. 980795.
- [12] Citron, S. J., O'Higgins, J. E., and Chen, L. Y., 1989, "Cylinder by Cylinder Engine Pressure and Pressure Torque Waveform Determination Utilizing Speed Fluctuations," SAE Technical Paper No. 890486.

- [13] Moro, D., Ponti, F., and Cavina, N., 1999, "In-Cylinder Pressure Reconstruction Based on Instantaneous Engine Speed Signal," *Proceedings of ICE Division of ASME, Fall Technical Conference*, Oct. 16–20, Ann Arbor, MI, ASME, New York.
- [14] Mock, R., and Meixner, H., "A Miniaturized High-Temperature Pressure Sensor for the Combustion Chamber of a Spark Ignition Engine," 1990–1991, *Sens. Actuators A25*(1–3), pp. 103–106
- [15] Poorman, T., Włodarczyk, M. T., Liangdao, X., Coleman, T., and Arnold, J., 1997, "Embedded Fiber-Optic Combustion-Pressure Sensors for Advanced Electronic Engine Control," *Proceedings Sensors Expo. Detroit*, Helmers Publishing & Expocon Management, Peterborough, NH, and Fairfield, CT.
- [16] Czarnocki, W. S., and Schuster, J. P., 1995, "Robust, Modular Integrated Pressure Sensor (for Automotive Application)," *Sensor95*, ACS Organisations GmbH, Wunstorf-Steinhude, Germany.
- [17] Gatowski, J. A., Balles, A. N., Chun, K. M., Nelson, F. E., Ekhian, J. A., and Heywood, J. B., 1984, "Heat Release Analysis of Engine Pressure Data," SAE Technical Paper No. 841359.
- [18] Heywood, J. B., *Internal Combustion Engine Fundamentals*, McGraw-Hill, New York.

Wear Assessment in a Biodiesel Fueled Compression Ignition Engine

A. K. Agarwal

Department of Mechanical Engineering,
Indian Institute of Technology, Kanpur,
Kanpur 208016, India
e-mail: Akag@iitk.ac.in

J. Bijwe

ITTEEC,
Indian Institute of Technology, Delhi,
New Delhi 110016, India

L. M. Das

Centre for Energy Studies,
Indian Institute of Technology, Delhi,
New Delhi 110016, India

Biodiesel is prepared using linseed oil and methanol by the process of transesterification. Use of linseed oil methyl ester (LOME) in a compression ignition engine was found to develop a highly compatible engine-fuel system with low emission characteristics. Two similar engines were operated using optimum biodiesel blend and mineral diesel oil, respectively. These were subjected to long-term endurance tests. Lubricating oil samples drawn from both engines after a fixed interval were subjected to elemental analysis. Quantification of various metal debris concentrations was done by atomic absorption spectroscopy (AAS). Wear metals were found to be about 30% lower for a biodiesel-operated engine system. Lubricating oil samples were also subjected to ferrography indicating lower wear debris concentrations for a biodiesel-operated engine. The additional lubricating property of LOME present in the fuel resulted in lower wear and improved life of moving components in a biodiesel-fuelled engine. However, this needed experimental verification and quantification. A series of experiments were thus conducted to compare the lubricity of various concentrations of LOME in biodiesel blends. Long duration tests were conducted using reciprocating motion in an SRV optimol wear tester to evaluate the coefficient of friction, specific wear rates, etc. The extent of damage, coefficient of friction, and specific wear rates decreased with increase in the percentage of LOME in the biodiesel blend. Scanning electron microscopy was conducted on the surfaces exposed to wear. The disk and pin using 20% biodiesel blend as the lubricating oil showed lesser damage compared to the one subjected to diesel oil as the lubricating fluid, confirming additional lubricity of biodiesel. [DOI: 10.1115/1.1501079]

Introduction

The world's economy depends upon the burning of fossil fuel equivalent of some 180 million barrels of oil each day. The consumption rate is equivalent to an annual burning of what nature took about one million years to accumulate as fossil deposits. The world at present is confronted with the twin crises of fossil fuel depletion and environmental degradation. Indiscriminate extraction and lavish consumption of fossil fuels have led to reduction in underground-based carbon resources. The search for an alternative fuel, which promises a harmonious correlation with sustainable development, energy conservation, management, efficiency, and environmental preservation, has become highly pronounced in the present context.

More than 6.5 million diesel engines are being used in the Indian agricultural sector for various activities. It is impossible to do away with these existing systems and hence alternative fuels are being expeditiously sought. As far as the application in rural agricultural sectors of a developing country like India is concerned, such internal combustion engines should preferably utilize alternative fuels of bio-origin, which are locally available. With agricultural conditions prevailing in India, the adaptation of the fuel to the existing engines has been more pronounced compared to developing a new engine for available alternative fuel. This allows the replacement of diesel fuel by vegetable oils, and for a short-term fuel as a blend.

Alternative fuels are generally referred as cleaner fuels in comparison with gasoline or diesel fuels. While using a new fuel, it is extremely important to strike a balance among many conflicting parameters involving not only the performance and emission char-

acteristics of the engine, but also the overall life of the system. The combustion-related properties of vegetable oils are somewhat similar to diesel oil ([1]). Neat vegetable oils or their blends with diesel pose various long-term problems in compression ignition engines, e.g., poor atomization characteristics, ring-sticking, injector-choking, injector deposits, injector pump failure, and lube oil dilution by crank-case polymerization. Such problems do not arise with short-term engine operations. Sometimes, the engine fails catastrophically, when operated on neat vegetable oils continuously for a longer period. The properties of vegetable oils such as high viscosity, low volatility, and polyunsaturated character are responsible for these problems ([2,3]). The deposits on the piston top are black, thick and hard, and at the edge, they fill the clearance in the top dead center position so that they will polish the bore at each stroke. A major problem using vegetable oil in engine operation is the deposits in the upper piston ring groove and on the piston ring. The piston ring gets stuck in the groove thereby weakening and decreasing the engine performance, as the combustion becomes erratic. The deposits grow, and the efficiency decreases. Another problem is that the vegetable oil and gases escapes through clearance into the engine crankcase as the sticking ring fails to seal adequately. As a result, it contaminates the lubrication oil, which leads to tough, rubber-like coating on the engine parts and the walls of the case, the fuel pump, and camshaft and push rods. The coatings on these parts can cause trouble or even breakdown. The build-up of carbon deposits is generally attributed to the large molecular size and resulting high viscosity of the medium-chain and long-chain triglycerides that constitute most commercial vegetable oils ([2,4]).

Keeping this in view, engine conditions were investigated using the newly developed biodiesel fuel. Transesterification is probably the most effective and widely used technique for formulating suitable fuel for CI engines from vegetable oils in order to avoid these problems. Transesterification was known as early as 1864, when Rochleder described glycerol preparation through ethanolysis of

Contributed by the Internal Combustion Engine Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the ICE Division, February 2001; final revision received by the ASME Headquarters, February 2002. Editor: D. N. Assanis.

castor oil. Transesterification is a chemical reaction that aims at substituting the glycerol of the glycerides with three molecules of mono-alcohols such as methanol thus leading to three molecules of methyl ester of the vegetable oil. The idea of chemically altering vegetable oils was noted even before World War II. Walton wrote in 1938 "to get the utmost value from vegetable oils as fuels, it is academically necessary to split the glycerides and to run on the residual fatty acid" because "the glycerides are likely to cause an excess of carbon in comparison" [5].

The stoichiometry of the overall transesterification reaction requires 1 mole of triglyceride for 3 moles of glycerol. This is a reversible reaction, which is acid or alkali catalyzed, involving stepwise conversions of triglycerides to diglycerides to monoglycerides to glycerol, thus producing 3 moles of ester in the process ([6]). A mixture of anhydrous alcohol and reagent (KOH) in desired proportions is combined with moisture-free vegetable oil. The materials are maintained at 65–70°C and allowed to settle by gravity for 24 hours. These esterified oils can be produced in a chemical laboratory by the above method. Some of the esters are commercially available. The molar ratio of the reagents, the presence of water, and free fatty acid determine the effectiveness of catalyst. Higher molar ratio (6:1) favors the use of potassium hydroxide (KOH) over potassium methoxide (KOCH_3), while lower molar ratio (3:1) favors the use of KOCH_3 . Alkaline catalysts have the advantage of being less corrosive to industrial equipment than acid catalysts. Transesterification breaks the triglycerides in vegetable oils to two components, namely fatty acid esters and glycerol. Glycerol is a valuable byproduct, which is used in pharmaceuticals, cosmetics, toothpaste, and many other commercial products. Biodiesel is often blended with petroleum diesel to offset its high production cost ([7,8]).

Background Experiments

For the present work, linseed oil methyl ester (LOME) was prepared using methanol in the presence of potassium hydroxide as catalyst. Glycerol is the byproduct of the esterification reaction. Use of LOME in compression ignition engines was found to develop a highly compatible engine-fuel system with low emission characteristics. The physical and combustion related properties of the fuels were determined in the laboratory after esterification and most of them were found to be close to that of diesel oil ([9]).

Performance and emission test was carried out on one of the engines, which had been subjected to preliminary run-in. This test was aimed at optimizing the concentration of ester in the biodiesel blends. To achieve this, several blends of varying concentrations were prepared ranging from 0% (Neat diesel oil) to 100% (Neat biodiesel) through 5%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, and 75%. These blends were then subjected to performance and emission test on the engine. Performance data were analyzed from the graphs recording power output, torque, specific fuel consumption, and smoke density for all the blends of biodiesel. 20% biodiesel was found to be the optimum blend based on maximum thermal efficiency and smoke opacity considerations ([2,9]). Two similar engines were subjected to a long-term endurance test for 512 hours. The engine system selected is a portable diesel engine of 4 kW rating with an alternator coupled to it. This unit, manufactured by Perry & Co., India, is a compact, portable 4 kW genset run by diesel fuel. This is a single cylinder, four-stroke, vertical, water-cooled system having a bore of 86 mm and a stroke of 110 mm. At the rated speed of 1500 rpm, the engine develops 4 kW power output in pure diesel mode. The compression ratio is 16.7. The inlet valve opens 4.5 deg BTDC and closes 35.5 deg ABDC. The exhaust valve opens 35.5 deg BBDC and closes 4.5 deg ATDC. The test engine is directly coupled to a 220 volts AC generator of sufficient capacity to absorb the maximum power produced by the engine. These engines were run for 32 cycles, each of 16 hours continuous running at rated speed and a fixed pattern of loading. The first engine operated on 20% biodiesel

blend, and the second engine on neat mineral diesel oil. The engines had similar history and were rebuilt with new liner, piston rings, bearings, etc. All metallic wear debris and corrosive wear debris generated during engine operation by wear of vital moving components were washed away by the lubricating oil and accumulated in the lubricating oil sump. Lube oil samples were drawn from the oil sump after intervals of every 128 hours of engine operation for conducting various tribological studies. Samples were taken after every 128 hours instead of industry standard of every 50 hours, in order to reduce the number of samples for the analysis. No major engine breakdown was observed during the entire range of engine operation of over 512 hours. The injector coking and deposition of carbon on cylinder head, piston top, was found to be significantly lower for biodiesel operated engine system ([10]). Since the engine used in this study are small engines, some of the fuel spray jets impinges upon liner walls in the combustion chamber and pass away to the oil sump along with the blow-by gases escaping through the piston rings. In the process, it affects the lubricating oil film present between the piston rings and liner wall. Though the engines are designed to minimize these losses, they can not be eliminated completely in small engines, thus leading to wear of various components.

Wear is defined as the progressive loss of substance from the operating surface of a body occurring as result of relative motion at the surface. It is a slow but a continuous process of removal of material from one or more elements. Wear depends on the load, speed, soot formation, temperature, hardness, surface finish, presence of foreign materials, corrosion and environmental conditions. During the process of wear, materials may deform or may be transferred from one element to another forming debris or the physical dimension may diminish. Wear particle analysis is a powerful technique for nonintrusive examination of the oil-wetted parts of a machine. The wear particles present in the lubricating oil provide a detailed information about the condition of the engine. A number of tests such as atomic absorption spectroscopy (AAS), ferrography, etc. were performed to estimate engine conditions indirectly through oil analysis. These tests provided information on the wear debris composition, wear debris size, type of wear, etc.

Atomic Absorption Spectroscopy

Atomic absorption spectroscopy was used for quantitative and qualitative analysis of wear debris of lubricating oils. The data was correlated with the extent of wear, the performance characteristics of either lubricating oils or diagnosis of failure of moving components. Atomic absorption spectroscopy (AAS) works on the principle of absorption interaction, where atoms in the vapor-state absorb radiation at a certain wavelength that are well defined and show the characteristics of a particular atomic element. In this process, the source of radiation projects a beam of a specific wavelength through a pure flame (air-acetylene) on to a sensor and the amount of radiation arriving at the photo sensor is recorded. The fluid sample is introduced into the flame and vaporized. The amount of radiation arriving at the photo sensor is reduced in proportion to the quantity of the specific element present in the sample.

The concentration of various metals present in the lubricating oil samples from 20% biodiesel and neat diesel-fuelled CI engines was evaluated to study the wear of different parts and material compatibility of the new fuel with the existing engines. Since many sliding components were involved, it was anticipated that the wear debris originating from different metallic parts appeared in the lubricating oil. Hence, various elements such as Fe, Cu, Zn, Cr, Mg, Co, and Pb were analyzed by following procedure.

Approximately 10 gms of lube oil was weighed in a silica crucible and burnt at 450°C for 4 hrs and at 650°C for 2 hrs. The residual ash was dissolved in concentrated HCl and the mixture was diluted with distilled water to make 100 ml of solution (Acid:

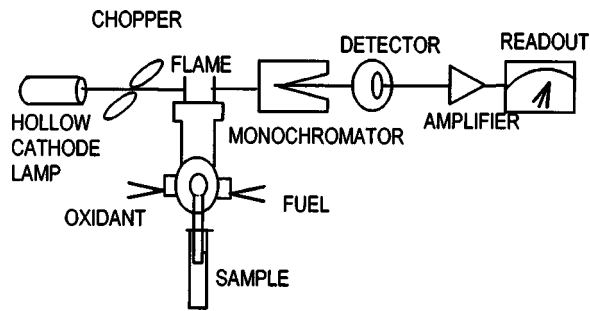


Fig. 1 Schematic diagram for atomic absorption spectroscopy

Water: 1:100) standard solution of various metals with concentration ranging from 5 PPM to 20 PPM were prepared and subjected to AAS. The results are shown in Figs. 1–8.[10]

The wear metal concentration seems to be typical for this type of engines. AAS results for various wear metals shown in Figs. 2–8 conclude that the lubricating oil drawn from biodiesel-operated engine has lower absolute amount of wear metals and lower or nearly same rate of increase in concentration of wear

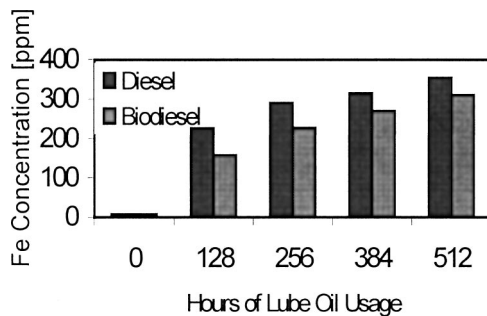


Fig. 2 Iron concentration as a function of lube oil usage

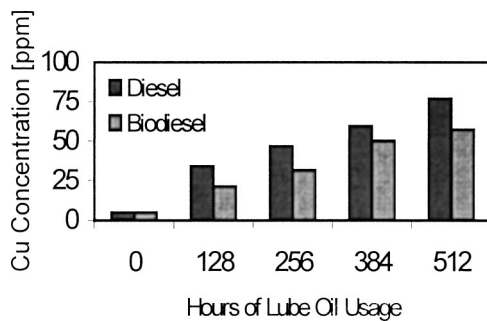


Fig. 3 Copper concentration as function of lube oil usage

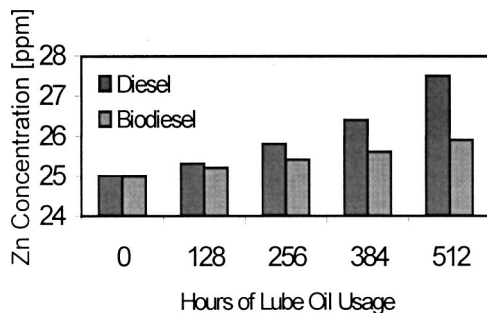


Fig. 4 Zinc concentration as a function of lube oil usage

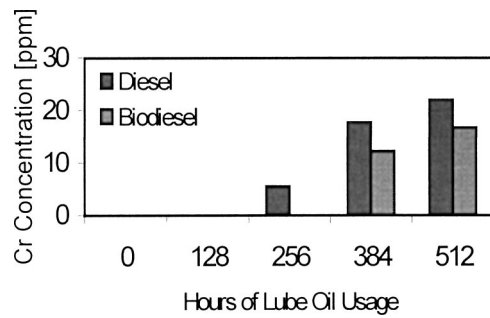


Fig. 5 Chromium concentration as a function of lube oil usage

metals. These results indicate that 20% amount of biodiesel present in significantly reduces wear of moving components.

To study the wear debris sizes, and concentration, ferrography was done on the lubricating oil samples drawn from both the engines after 128 and 512 hours of engine operation.

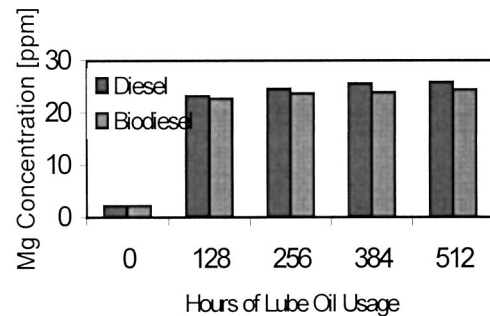


Fig. 6 Magnesium concentration as a function of lube oil usage

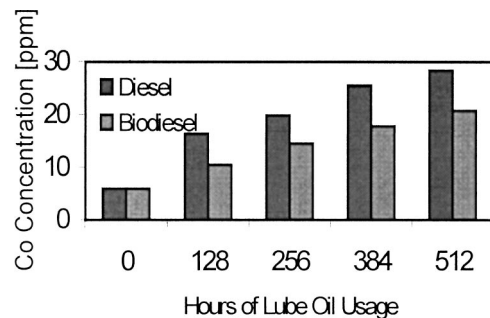


Fig. 7 Cobalt concentration as a function of hours of fuel usage

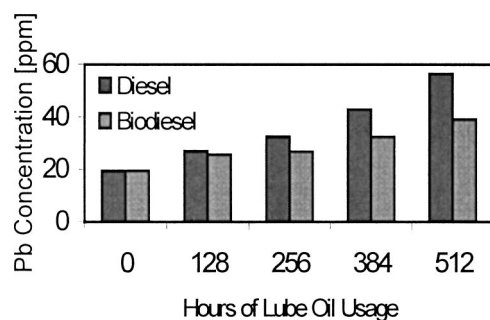


Fig. 8 Lead concentration as function of hours of lube oil usage

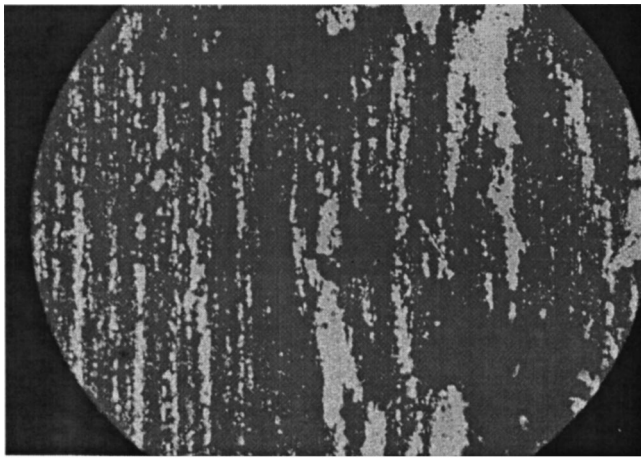


Fig. 9 Ferrogram of the lube oil drawn from 20% biodiesel operated engine after 128 hours of operation

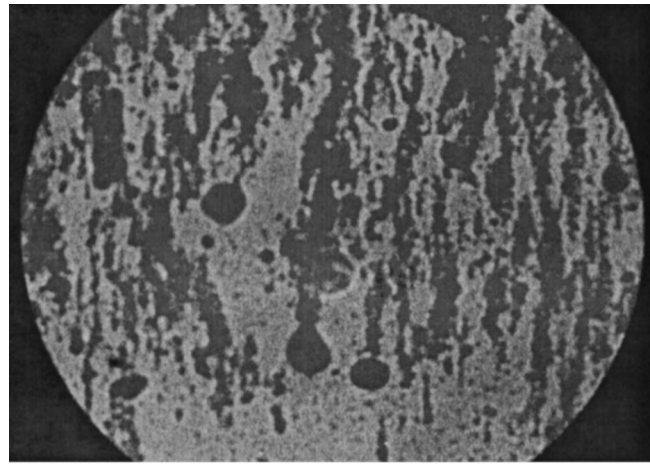


Fig. 11 Ferrogram of the lube oil drawn from 20% biodiesel operated engine after 512 hours of operation

Ferrography

During the life cycle of an IC engine, microscopic particles wear away from various moving parts. The size, distribution, number, and shape of the microscopic particle can provide information about the condition of the part they originate from, and the form of wear and wear mechanisms. Wear particle analyses in lube oil can therefore be used as a tool for monitoring not only the health of the engine oil but of the moving parts as well. Lube oil samples were subjected to ferrography to investigate the wear debris size and concentration in the oil samples. DR ferrograph by Fox Boro/Trans-Sonics Inc. was used for this study.

In the ferrogram, opaque objects appeared black, transparent objects appeared light gray and translucent materials appeared gray. Ferrography was done on the basis of classification of wear debris by color, texture, shape, concentration, and composition. Ferrography is a two-step process. The first step was of slide preparation and second step was of particle analysis. Bi-chrome microscope was used for particle analyses. Origin of the particle could be inferred from the composition of the particles, e.g., indications of bronze, copper or a special type of steel may be traced to specific bearings. Quantification of the ratio of sizes of large and small debris was not possible due to unavailability of dL/dS ratio sensor.

It is evident from Fig. 9 and 10 that the concentration of me-

tallic wear debris is visibly lower for biodiesel-operated engines. Similarly, the ferrograms for 20% biodiesel and diesel-operated engines after 512 hours of engine operation in Fig. 11 and Fig. 12 indicate that the size of wear debris is smaller for biodiesel-operated engines. The type of wear identified is sliding wear, which is in conformity with the sliding motion of various engine parts. These results are in complete agreement with the AAS results, suggesting additional inherent lubricating properties for biodiesel present in fuel.

The different wear pattern can be due to the following reasons. The engines used in this study are small engines, and the engine walls possibly get some amount of fuel spray droplets in every injection cycle. These fuel droplets mix with the lubricating oil, which adheres to the cylinder liner wall surfaces, diluting it to some extent. The diluted lube oil gets swiped back into the crankcase due to the dynamics of the piston rings. In the process, the diluted lube oil acts as a lubricant between the cylinder liner and piston rings. It is expected that the lubricating properties of biodiesel and diesel fuel's play a predominant role at such a juncture and there is a need to investigate and quantify the lubricity of these fuels in order to find the difference in wear patterns. To investigate this explanation further, friction and wear patterns of

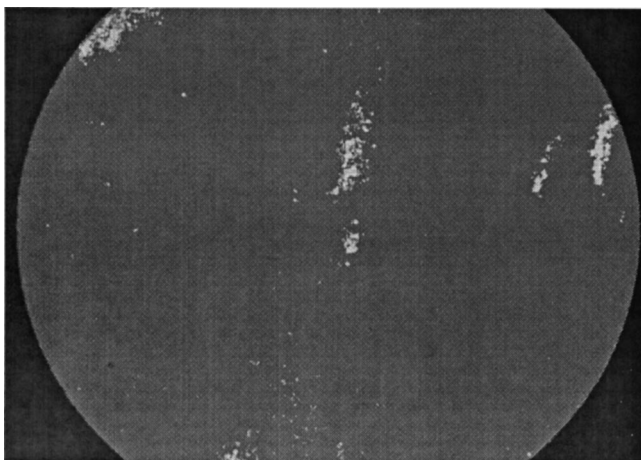


Fig. 10 Ferrogram of the lube oil drawn from diesel operated engine after 128 hours of operation

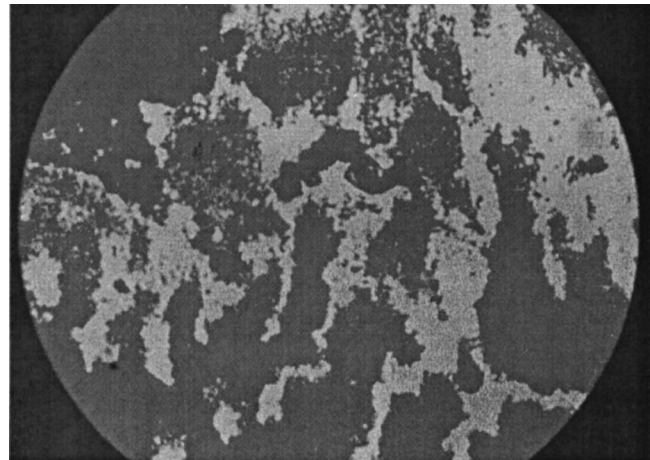


Fig. 12 Ferrogram of the lube oil drawn from diesel operated engine after 512 hours of operation

various combinations of these fuels were used as lubricating fluids in a SRV Optimal wear tester, which simulates the same type of motion as a real life diesel engine.

Experiments on Wear Tester

The SRV optimal wear tester was used for studying wear between rubbing surfaces. The machine (Fig. 13) used in this study allows a small section of piston ring to be moved relative to a small disk, made from the liner material. The device is such that the stroke of the ring piece and frequency can be varied to control the speed at which the ring piece moves. A force, that can be adjusted, pushes down the fixture holding the ring piece. Heating can control the temperature of the liner disk. Though, the machine is a wear tester, it does not completely match the characteristics of the piston ring in an engine, and also due to physical restriction, the ability to change the stroke on the tester was limited. However, fairly good results for the lubricity and specific wear rates of the surfaces for different lubricating fluids were achieved. For the present set of investigations, six pins were made from the top compression ring of the piston. The top ring was the hardest piston ring with hard chrome plating. Six disks were machined from the engine liner made of high phosphorous cast iron. A pin was held tightly in an aluminum holder provided in the SRV tester. The pin is slid in reciprocating motion against the disk made from the engine liner dipped in the lubricating fluid. Four different blends of LOME (0%, 20%, 40%, and 60%) with diesel oil were investigated. The following set of parameters given below was studied.

stroke length	1 mm
oscillation frequency	50 Hz
load	500 N
duration of runs	4 Hours

The engine used in the long duration tests has a stroke of 86 mm and operates at 1500 RPM with an oscillating frequency of 25 Hz. This can be noticed here that at this stroke and frequency, the maximum speed of the ring in the engine is much higher than in the wear tester. Lower speeds in the wear tester result in thin films of lubricating oil on the liner wall. In the fired engine used, the peak combustion chamber pressures that push on the back of the top ring can reach upto 9.4 MPa. This pressure, when applied to the ring piece, would be equivalent to 415 N. Since the speed of

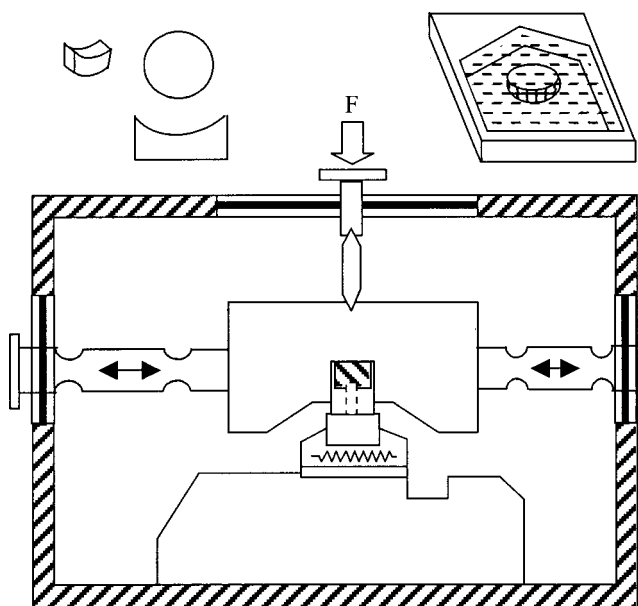


Fig. 13 SRV optimal wear tester, pin and disk arrangement

Blend Concentration vs. Coefficient of Friction

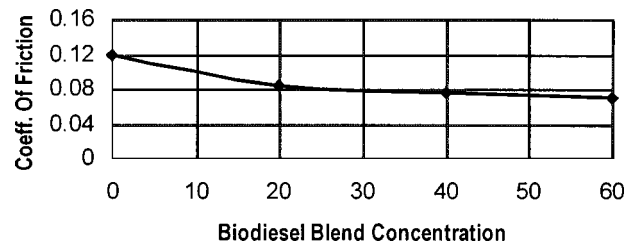


Fig. 14 Coefficient of friction of biodiesel blend concentration

the pin was small, it was decided to go for higher load (500 N). It should also be noted that the peak pressure in the four-stroke engine does not last over the entire cycle of the engine. High pressures are present only for a part of each cycle. During low pressure in the cycle, only the ring tension exerts a force on the liner. This is approximately equivalent to a 14 N load on the ring piece ([11]). In the wear tester, the loading remains constant, and hence all the engine conditions are not simulated because of the constraints of the machine.

In this study, friction coefficient was monitored continuously and then an average taken. The wear of pin and disk was measured by weighing the components before and after the tests. All the components were cleaned in ultrasonic bath with acetone, before and after the test, in order to remove foreign particle deposits.

Specific wear rates for the pin and the disk were calculated by the equation given below.

$$K_o = V/(LD)m^3/Nm \quad (1)$$

where

- K_o = specific wear rate,
- V = volume of the worn out material (m^3),
- L = load in Newton and
- D = total sliding distance (m).

The volume of worn out material is calculated using weight of worn out material and its density. The distance slid can be calculated by Eq. (2).

$$D = 2A \nu t \quad (2)$$

where

- A = Stroke length (m),
- ν = oscillation frequency (Hz), and
- t = the time period of run in seconds.

Efforts were also made to conduct the tests in 80% and 100% LOME blends but it was noticed that there is a sudden increase in noise after running the machine for more than an hour. The LOME gets completely blackened and thick due to oxidation and the oxidized particles formed increase the coefficient of friction in the contact zone. The coefficient of friction abruptly rises from the initial value (0.04). Hence, tests for higher percentages of LOME were abandoned. These results suggest that biodiesel leads to higher oxidation of the fuel, hence anti-oxidant additives are required in the biodiesel fuel to save it from deterioration with time.

The test results from the four lubricating fluid samples are represented in the following paragraphs. Figure 14 represents the average friction coefficient as a function of percentage of LOME in lubricating fluids. The figure shows that with increase in percentage of LOME in the lubricating fluid from 0 to 60, the coefficient of friction has continuously dropped from 0.12 to 0.07.

Figure 15 represents the specific wear rate for piston and disk as a function of percentage of LOME in the lubricating fluids. This figure also shows some interesting results in conformity with earlier observations. Specific wear rates for both pin and disk decreased with the increase in percentage of LOME in lubricating

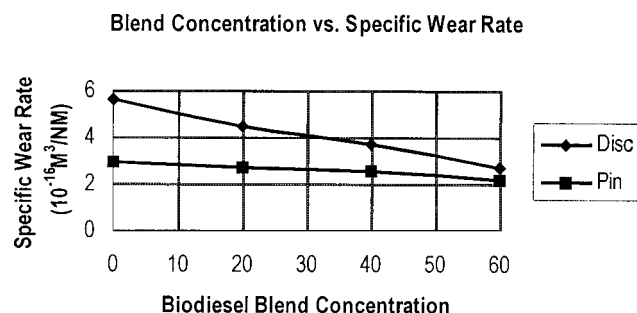


Fig. 15 Specific wear rates for disk and pin as a function of biodiesel blend concentration

fluids. Specific wear rates for disk submerged in diesel were maximum ($5.6 \times 10^{-16} \text{ m}^3/\text{Nm}$) and they were lowest for 60% LOME blend ($2.7 \times 10^{-16} \text{ m}^3/\text{Nm}$) with similar trends for pin material. Specific wear rates of disk material were higher than that for pin material because pin material was almost three times harder than the disk material. These tests clearly indicates that LOME has inherent lubricity properties, and this leads to lower wear of moving engine parts and thus longer life for the engine.

These disks were also subjected to scanning electronic microscopy. SEM micrographs show the surface topography with high magnifications.

Scanning Electron Microscopy

SEM was conducted on two disks, one subjected to neat diesel oil and the second subjected to 20% biodiesel blend as lubricating fluid. The micrographs for the two disks with different magnifications are shown in Fig. 16–19.

Micrographs 16 and 17, when compared to the micrographs 18 and 19, which are at the same magnifications indicate the difference in topography. In the case of disk subjected to 20% biodiesel blend, the overall surface appeared to be significantly smoother than the disk subjected to diesel indicating the reduction in wear. This also supports the fact that biodiesel is a better lubricant since it reduces the wear of the disk. The portion on the right side of micrographs 16 and 18 shows unworn surface of the disk when compared to the left side of the micrograph. High extent of grooving can be seen in the case of diesel-disk, which is due to body abrasive wear. The initial wear debris from the sliding surfaces get trapped in the contact zone and keep on abrading the surface till it is thrown out of the contact zone. Micrographs 17 and 19 (same

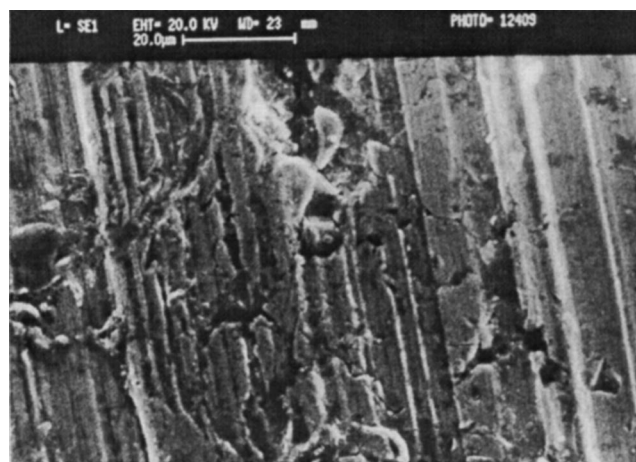


Fig. 17 High magnification micrograph of the disk subjected to diesel as lubricating fluid

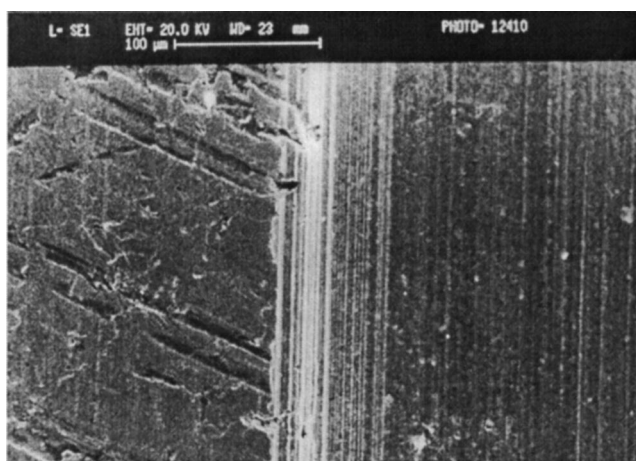


Fig. 18 Micrograph of the disk subjected to 20% biodiesel as lubricating fluid



Fig. 16 Micrograph of the disk subjected to diesel as lubricating fluid

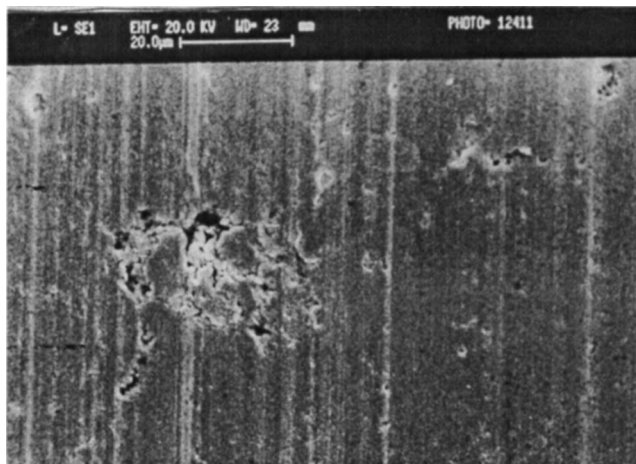


Fig. 19 High magnification micrograph of the disk subjected to 20% biodiesel as lubricating fluid

magnification) indicate a drastic difference in the topography of the surfaces. Disk surface subjected to diesel is extensively abraded at various places. It is evident from the above study that the self-lubricity of biodiesel results in reduced wear of engine parts.

Conclusions

Biodiesel can be readily adopted as a partial substitute for diesel in today's developing economies of the world. Transesterification is an appropriate method for the modification of vegetable oils, thus producing a suitable diesel fuel. Biodiesel is an environment-friendly fuel as it can be successfully used for long duration as none of the operational or durability problems were observed.

Atomic absorption spectroscopy tests conducted on lubricating oil samples drawn from biodiesel-operated engines suggested low concentration of wear metals. Ferrography tests showed smaller size and lower concentration of wear debris for biodiesel-operated engines. Though increase in wear debris concentrations was observed with usage of both biodiesel and diesel operated engine systems.

Additional tests conducted on SRV optimal wear tester to simulate the engine conditions partially for quantification of lubricity reflected that the lubricating fluids with biodiesel show lower coefficient of frictional low specific wear rates for moving components of the engines.

The scanning electron microscopy conducted on the disks subjected to wear tester reflected that biodiesel fuel is less damaging to the moving parts. These tests confirmed inherent additional lubricating properties of the LOME, i.e., biodiesel. Thus, it is apparently proved from the present investigation that biodiesel is not only a good choice for environmental reasons but it also keeps the engine in a better health, improving the life of its vital moving components.

Thus biodiesel fuel has proved to be a competent, engine friendly and environment friendly substitute for the mineral diesel oil.

Acknowledgments

The authors would like to express their gratitude towards Shri. Attar Singh, Engines and Fuels Laboratory, Center for Energy Studies, Indian Institute of Technology, Delhi, for his encouragement and constant guidance in conducting the tortuous series of experiments. The authors also thank the Council for Scientific and Industrial Research, New Delhi, India, for granting the fellowship for conducting research on this topic.

References

- [1] Agarwal, Avinash Kumar, 1999, "Performance Evaluation and Tribological Studies on a Biodiesel-Fuelled Compression Ignition Engine," Ph.D. thesis, Center for Energy Studies, Indian Institute of Technology, Delhi, India.
- [2] Krawczyk, Tom, 1996, "Biodiesel: Alternative Fuel Makes Inroads but Hurdles Remain," *Inform*, No. 8, 7(8), pp. 801–814.
- [3] Teri Energy Economy Simulation and Evaluation Model, 1998, "Asia Least-cost Greenhouse Gas Abatement Strategy (ALGAS)," Tata Energy Research Institute, New Delhi.
- [4] Peterson, C. L., Auld, D. L., Bettis, B. L., Hawley, K. N., and Wagner, G. L., 1981, "Safflower and Rapeseed Oils as Fuel," *Alcohol and Vegetable Oil as Alternative Fuels*, Raleigh, NC, Apr. 7–9; Sacramento, CA, Apr. 21–23; Peoria, IL, Apr. 28–30.
- [5] Van Dyne, D. L., Weber, J. A., and Braschler, C. H., 1996, "Macroeconomic Effects of a Community Based Biodiesel Production System," *Bioresour. Technol.*, **56**, pp. 1–6.
- [6] Freedman, Bernard, Butterfield, Royden O., and Pryde, Everett H., 1986, "Transesterification Kinetics of Soybean Oil," *JAACS*, **63**(10), Oct.
- [7] American Biofuel Association, 1994, *Biodiesel: A Technology, Performance and Regulatory Overview*, Information Resources, Washington, DC.
- [8] Weber, J. A., and Johannes, K., 1996, "Biodiesel Market Opportunity and Potential Barriers," *Liquid Fuels and Industrial Products From Renewable Resources*, Proceedings of the Third fuel Conference, J. S. Wnditt, E. E. Gavett, C. Hansen, C. Peterson, M. A. Sanderson, H. Shapouri, and D. A. Van Dyne, eds., Sep. 15–17, Nashville, TN, ASAE.
- [9] Agarwal, Avinash Kumar, and Das, L. M., 2001, "Biodiesel Development and Characterization for Use as a Fuel in Compression Ignition Engine," *ASME J. Eng. Gas Turbines Power*, **123**, pp. 440–447.
- [10] Agarwal, Avinash Kumar, Bijwe, Jayashree, and Das, L. M., 2000, "Effect of Biodiesel Utilization on Wear of Vital Parts in Compression Ignition Engine," *Large Bore Engine Designs, Natural Gas Engines, and Alternative Fuels*, ASME, New York, ICE Vol. -35-2, pp. 45–54.
- [11] Richardson, Dana E., and Borman, Gary L., 1999, "Using Fiber Optics and Laser Fluorescence for Measuring Thin Oil Films With Application to Engines," SAE Paper No. 912388.

Combustion in Gas Fueled Compression: Ignition Engines of the Dual Fuel Type

G. A. Karim

Department of Mechanical and Manufacturing
Engineering,
University of Calgary,
Calgary T2N 1N4, Canada
e-mail: karim@enme.ucalgary.ca

In the dual fuel engine much of the energy release comes from the combustion of the gaseous fuel while only a small amount of diesel liquid fuel provides ignition through timed cylinder injection. Such operation with optimum conversion methods has the potential to provide operational characteristics that are comparable or superior to those of the corresponding diesel or spark ignition engines. These characteristics may be realized only if sufficiently effective measures can be ensured both for the avoidance of knock, usually at high loads, and incomplete gaseous fuel utilization at relatively light loads. An objective of this contribution is to demonstrate that the main effort needed to overcome the problems associated with the operation of gas fueled dual fuel engines is via a better control of the relatively complex processes of combustion. Both experimental and analytical modeling procedures for effecting optimum improvement to the combustion process are described. [DOI: 10.1115/1.1581894]

Introduction

The conventional internal combustion engine or clones of its mode of working will be with us for quite sometime to come. However, the demand for higher output efficiencies, greater mean effective pressures, increased reliability and ever-reduced emissions will continue to rise in intensity at the same time. Also, there is a perceived increased need, whether for environmental, economic or resource reasons to operate on a multitude of natural and processed gaseous fuels. Of course, there is a great research and development expended worldwide which will continue to provide solutions and make incremental progress toward achieving these goals. It is evident that the main effort needed is via a better control of the relatively complex processes of combustion and in particular its chemical aspects.

Much of the remedial technologies that have been successfully developed in recent years, which include catalytic conversion of exhaust gases and exhaust gas recirculation, were the product of empiricism of modifying the chemical reactions of flame propagation and of processing chemically the exhaust gases before discharging them into the atmosphere. Suitable controls of the physical aspects of the combustion process in engines, such as through improvements to the flow patterns, mixing and heat transfer have also contributed to the success of the internal combustion engine in recent years. However, the projected rate of progress in these areas is insufficient to fulfil the demanding objectives. There is a need to investigate various approaches that can provide a new dimension for manipulation the combustion process, especially its chemical aspects, so as to secure substantial progress towards improved efficiency, a better control of exhaust emissions and permitting a greater flexibility in choosing the fuel.

The Gas Fueled Dual Fuel Engine

In the dual fuel engine much of the energy release comes from the combustion of the gaseous fuel while a small amount of diesel liquid fuel provides ignition through timed cylinder injection. Such operation, with optimum conversion methods, has the potential to provide operational characteristics that are comparable or superior to those of the conventional diesel or spark ignition en-

gines. This may be achieved while displaying enhanced emission characteristics, quieter and smoother operation, improved low ambient temperature operation, and reduced thermal loading, [1]. This superior performance may be achieved only if sufficiently effective measures can be ensured both for the avoidance of knock, usually at high loads and incomplete gaseous fuel utilisation at relatively light loads. Usually, one of the main aims is to maximize the replacement of the diesel fuel by the gaseous fuel while maintaining acceptable levels of emissions and engine performance.

Dual fuel engine operation can be viewed in two broad categories depending on the relative amount of gaseous fuel to that of the diesel liquid fuel. The first more common category is to inject a relatively small quantity of diesel liquid fuel primarily to provide ignition of the lean mixtures of the gaseous fuel and air. The bulk of the energy release comes from the gaseous fuel components. The second category is associated with the addition of some gaseous fuel to the air of a fully operational diesel, well beyond the light load operational range. Ideally, there is a need for optimum variation in the liquid fuel quality in relation to the gaseous fuel supply so as to provide for any specific engine the best performance over the whole load range for any gaseous fuel, [2].

The high compression ratio diesel engine is imminently suitable for dual fuel operation with methane. However, largely by virtue of the provision of an adequate and consistently timed pilot liquid fuel injection and ignition, almost any gaseous fuel or vapor can be utilised to a varying degree of success in such engines. Gaseous fuels such as ethane, propane, butane, hydrogen, and ethylene have been employed. Moreover, various gaseous fuel mixtures, including those containing diluent gases such as natural gases, bio-gases, landfill gases, liquefied petroleum gases, etc., have been used. Also, information exists in the open literature relating to the performance of dual fuel engines when a wide range of liquid fuels was fumigated with the incoming engine air, [3].

The dual fuel engine has found applications in numerous fields. They have been employed in mobile sources such as fleet vehicles, heavy-duty trucks, buses, railway locomotives, and marine vessels, and in construction and agricultural field applications. Stationary applications have included engine electric power generators, pumps, and cogeneration sets, [4].

The movement to convert common diesel engines to operation on a wide range of gaseous fuels so far has tended to undermine

Contributed by the Internal Combustion Engine Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the ICE Division November 2001; final revision received by the ASME Headquarters July 2002. Associate Editor: D. Assanis.

some of the superior characteristics of the compression ignition mode. Some factors contributing to this trend are the following:

- i. Diesel engines tend to vary widely in type, size, and speed ranges, making their conversion too demanding while maintaining optimal performance.
- ii. The operation of dual fuel engines, especially when using some gaseous fuel mixtures, is severely limited by the onset of knock. Its avoidance while maintaining the efficient high compression ratios of diesel engines remains often too difficult to achieve satisfactorily.
- iii. There is often a need to develop operational methods to reduce the emission of oxides of nitrogen by a judicious choice of pilot fuel injection and gaseous fuel combustion characteristics.
- iv. Dual fuel engines at light load employ very lean gaseous fuel—air mixtures, which can lead to significant emissions of unconverted gaseous fuel, carbon monoxide and other products. Exhaust gas treatment catalytically tends to be less effective with some gaseous fuel mixtures.
- v. At present, the minimization of the liquid diesel fuel component relative to the gaseous fuel may increase exhaust emissions and enhance the incidence of knock and cyclic variations.
- vi. The interaction between the liquid pilot spray and the gaseous fuel is not only thermal but also has a kinetic component. These can bring about an extension to the ignition delay and increased emissions.
- vii. The direct injection of the gaseous fuel into the cylinder represents an added complexity, since a high pressure supply is needed and the mixing processes of the injected fuel with the pilot require careful control and understanding, [5].

A number of recent developments in our knowledge are being exploited not only to bring about improvements to our understanding of the processes of combustion but also to permit the development of new approaches that may be superior to those available at present.

Dual Fuel Combustion

The combustion process in a typical dual fuel engine depends both on the spray and ignition characteristics of the diesel pilot and on the type of gaseous fuel used and its overall concentration in the cylinder charge. The combustion energy release characteristics in a typical dual fuel engine reflect throughout the relatively complex physical and chemical interactions that take place between the combustion processes of the two fuel systems.

During the compression process, the premixed gaseous fuel-air charge becomes subjected increasingly with time to higher temperatures and pressures as the top dead center position is approached. The pre-ignition reaction activity within the cylinder charge may progress significantly, especially with certain reactive fuels and operating conditions. It can be shown that even with methane operation, some reactions can proceed during the compression process, even in the absence of pilot injection, [6]. As was shown through computational modeling employing detailed chemical kinetics for the oxidation reactions of methane in air, the mean temperature of the charge, as shown in Fig. 1, can increase substantially as the top dead center position is approached due to increasing chemical activity, [7]. This is especially prominent with mixtures that are not very lean. This increase in temperature may compensate more than for the lowering of the charge temperature due to changes in the thermodynamic properties of the charge, as a result of increased gas admission, as well as heat transfer effects. These reactions even in the absence of a pilot may produce some partial oxidation products such as radicals, aldehydes, and carbon monoxide. These can build up to significant amounts during the latter part of the compression stroke to influence the ignition and combustion processes of the pilot. These processes will contribute directly to the combustion of some of the gaseous charge en-

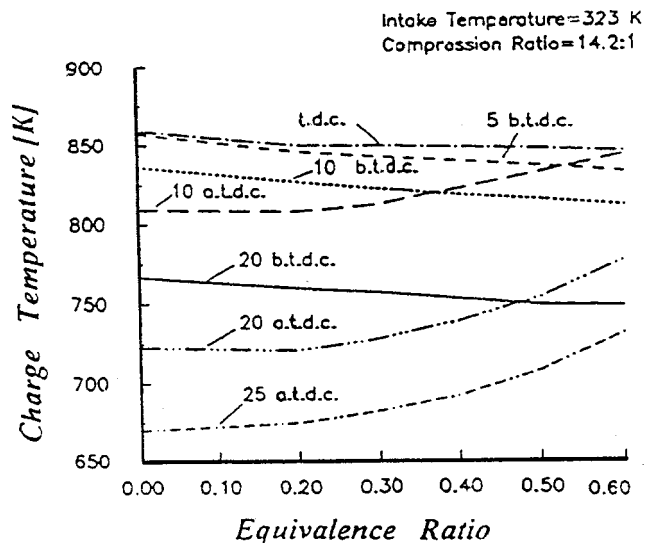


Fig. 1 Variations in the calculated temperature of a methane-air charge at different stages of compression and expansion in the absence pilot injection, with changes in equivalence ratio, $T_i = 323$ K

trained into the pilot and in its immediate vicinity and indirectly to the rest of the charge. Following pilot ignition, the energy release later on reflects that. However, turbulent flame propagation from the pilot ignition regions will not proceed throughout the charge until the concentration of the gaseous fuel is beyond a limiting concentration that would vary with the fuel employed and operating conditions, [8].

The combustion energy release rate of a dual fuel engine can be considered for convenience as made up of essentially three overlapping components. The first, (I), as shown in Fig. 2, is due to the combustion of the pilot. The second, (II), is due to the combustion of the gaseous fuel component that is in the immediate vicinity of the ignition and combustion centres of the pilot. The third, (III), is due to any preignition reaction activity and subsequent turbulent flame propagation (and sometimes autoignition) within the overall lean mixture.

With very lean gaseous fuel-air mixtures the bulk of the energy release comes from the ignition and subsequent rapid combustion of the small pilot zone (I). It comes also from the combustion of part of the gaseous fuel-air mixture entrained into the burning pilot and from the immediate surroundings of such a zone where higher temperatures and relatively richer mixture regions are present. As shown schematically in Fig. 3, under these conditions

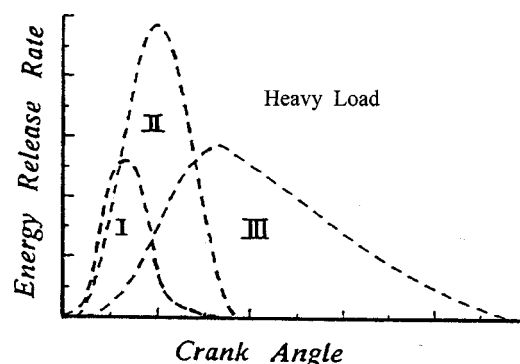


Fig. 2 Schematic representation of the different components of the combustion energy release rate in a dual fuel engine for a heavy load condition

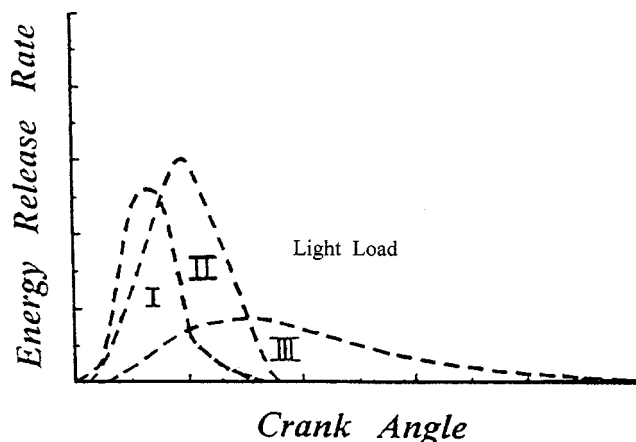


Fig. 3 Schematic representation of the different components of combustion energy release rate in a dual fuel engine for a light load condition

only relatively little contribution to the energy release may come from the bulk of the gaseous fuel-air charge further away from the pilot zone. Within the very lean mixtures, no consistent flame propagation will take place from the ignition centres and pilot influenced burning regions. An increase in the size of the pilot zone whether through increasing the mass of fuel injected or its distribution in very lean mixtures will tend to increase more than proportionally the total energy released and its associated rates. Greater amounts of gas-air mixtures then will be oxidized due to the larger amount of mixtures entrained within the pilot combustion zone and as a result of the widening burning regions in their vicinity, [9]. Greater energy release and rates will be evident also due to some partial flame propagation and increased pre-ignition reaction activity of the rest of the charge, [10].

Increasing the concentration of the gaseous fuel further will permit eventually after pilot ignition, flame propagation throughout the rest of the charge resulting, as shown in Fig. 4, in a sudden increased contribution to the total energy release. A continued

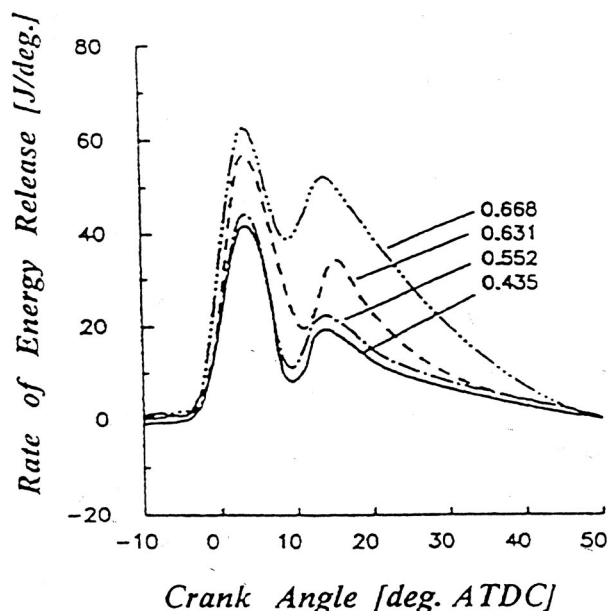


Fig. 4 Variations in the effective energy release rates with crank angle for different total equivalence ratios at a constant pilot quantity of 0.227 kg/h, for methane operation

increase in the concentration of the gaseous fuel in air will result in a greater overlap between the second and third energy release regions and will lead to their amalgamation, further increasing much of the energy release immediately following the commencement of the auto-ignition of the pilot.

The Ignition Delay

In the homogeneous gas fueled dual fuel engine the ignition delay, as shown typically in Fig. 5, displays trends significantly different from those observed in the corresponding diesel engine operation. The delay tends to increase with the increased gaseous fuel admission up to a detectable maximum value and then drops to a minimum well before reaching the total stoichiometric ratio (i.e., based on the combined gaseous and liquid fuels and the available air). The admission of propane, which is regarded normally as a relatively fast reactive fuel in comparison to methane, increases the delay to values greater than those observed with methane.

The introduction of the gaseous fuel with the intake air produces variations in the physical and transport properties of the mixture, such as the specific heat ratio and to a lesser extent the heat transfer parameters. Also, changes in the intake partial pressure of oxygen due to air displacement by the gaseous fuel, changes in the pre-ignition reaction activity and its associated energy release and the effects of residual gas can bring about substantial changes to the pre-ignition processes of the pilot and hence the length of the delay period.

For any constant compression ratio, unthrottled engine operating at a fixed intake temperature, in the absence of pilot fuel injection, the peak temperature at TDC varies markedly with the type of gaseous fuel used and its concentration in the cylinder charge. As shown in Fig. 6, the addition of hydrogen to the engine intake air reduces the maximum temperature level of the charge only relatively little with the increase of equivalence ratio. The corresponding values of temperature with the addition of methane or propane decrease more markedly. With propane addition for the case shown, a drop in temperature of around 100 K can be observed for the stoichiometric mixture. The heat loss during compression, which is also strongly dependent on the gaseous fuel used and its concentration, influenced the temperature also. As shown in Fig. 7, the corresponding calculated maximum heat

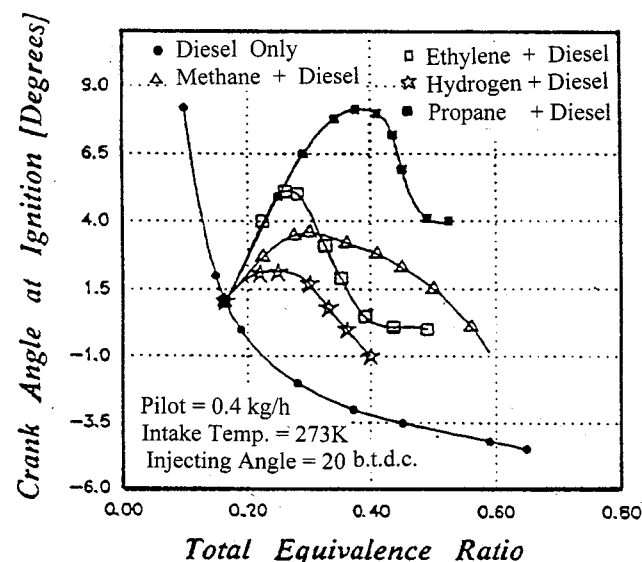


Fig. 5 Variations of the ignition point with total equivalence ratio for different fuels at constant pilot quantity. The corresponding diesel operation is shown.

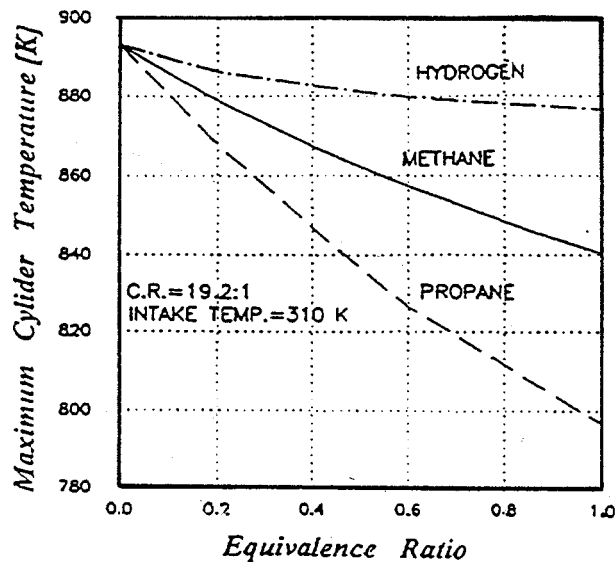


Fig. 6 The variation of the calculated temperature at TDC with equivalence ratio in the absence of pilot injection for three different fuels

transfer rates decreased only a little with the increased admission of propane or methane, but the rates increased very significantly with the increased admission of hydrogen, [7].

The resulting corresponding variations of the maximum pre-ignition energy release rates with changes in charge equivalence ratio in the absence of pilot fuel ignition is shown in Fig. 8. It can be seen that the contribution of the pre-ignition energy release to the development of temperature is also strongly dependent on the gaseous fuel used. The increased admission of hydrogen continued to increase the energy release rate, which has a positive contribution to enhance the charge temperature. With methane or propane admission the peak values of the energy release rate display maximum values that are on the lean side of the stoichiometric. For any equivalence ratio, the rates associated with methane op-

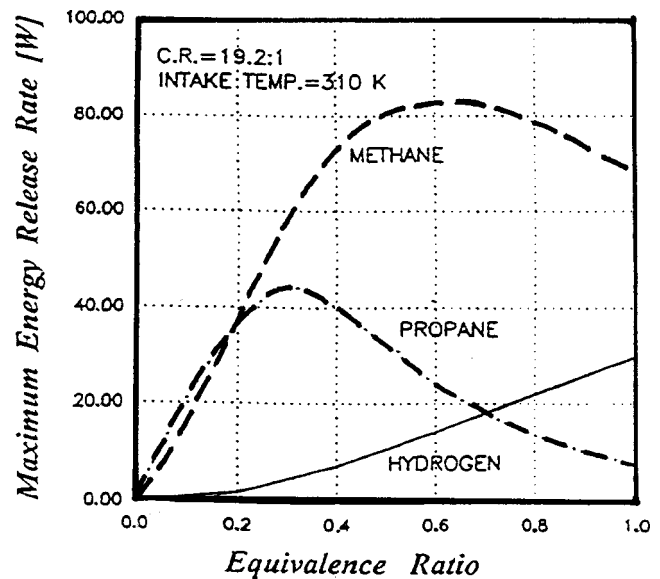


Fig. 8 Variation of the calculated maximum energy release rate with equivalence ratio in the absence of pilot injection for propane, methane, and hydrogen operation

eration are higher than with propane or hydrogen. This is somewhat unexpected. It is a reflection of the changes to the mean charge temperature with fuel gas admission. The peak value of the rate for propane operation is obtained with a leaner mixture than that with methane. For the case shown, the two equivalence ratios are around 0.30 and 0.60 for propane and methane operation, respectively. Hence, the increased admission of either methane or propane may not necessarily enhance the pre-ignition reactivity of the gaseous fuel-air mixture under engine conditions.

The admission of diluents such as carbon dioxide or nitrogen into a diesel engine charge also can produce changes in the compression temperature. As shown in Fig. 9, the mean value of the charge temperature decreases virtually linearly with increasing the concentrations of carbon dioxide, while the admission of nitrogen hardly affects the temperature. These trends result primarily from the marked changes in the effective specific heat ratio of the mix-

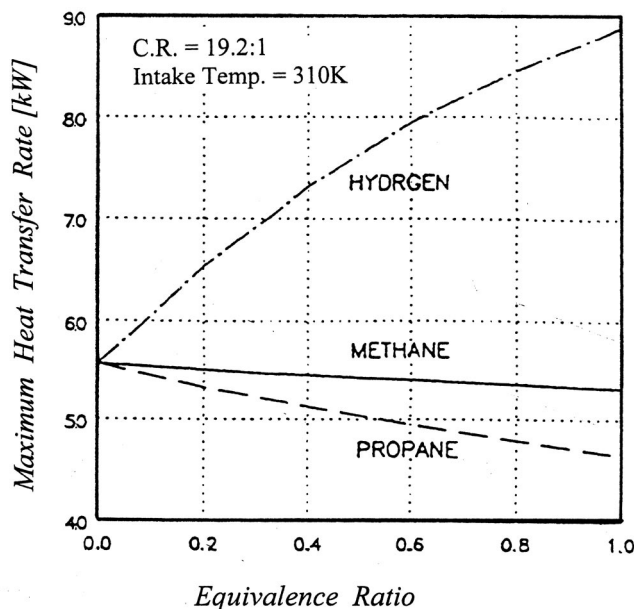


Fig. 7 Variations in the calculated maximum heat transfer rate with equivalence ratio in the absence of pilot fuel for hydrogen, methane, and propane admissions

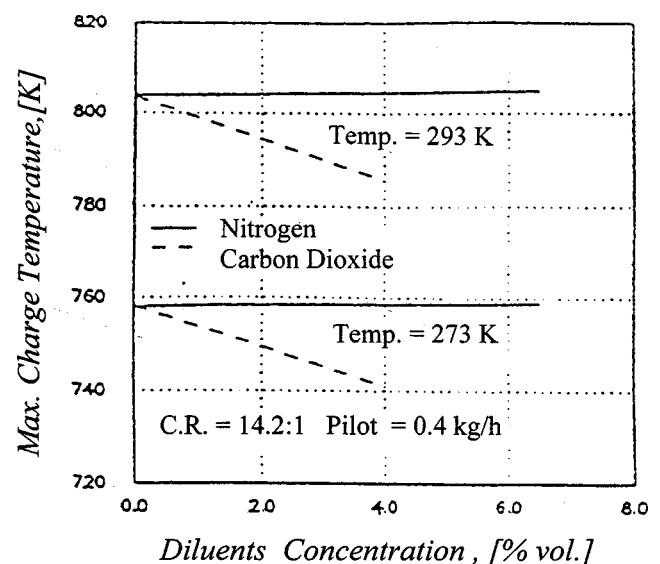


Fig. 9 Variations of the calculated charge temperature at ignition with the admission of nitrogen and carbon dioxide in a diesel engine with a fixed pilot quantity

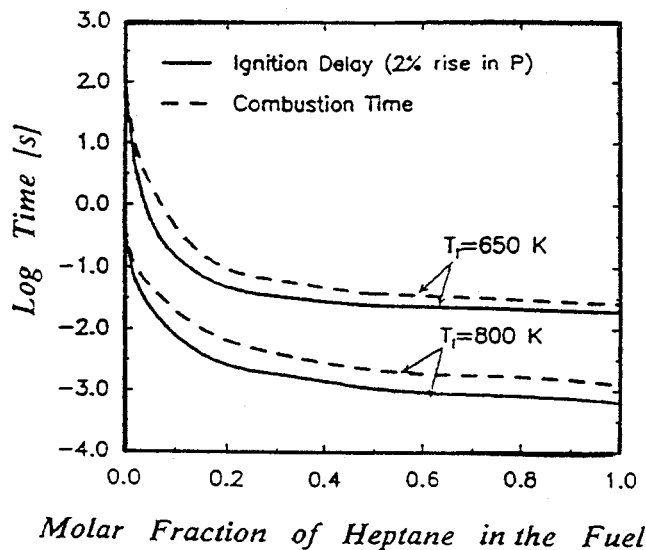


Fig. 10 Variations of the calculated ignition delay and total combustion times with the molar fraction of n-heptane in heptane-methane stoichiometric mixtures in air for two initial temperatures under adiabatic constant volume conditions, $P_i = 2.8$ MPa

ture with the increased presence of carbon dioxide. Also, it would be expected that the increase in the length of the delay with the increased admission of carbon dioxide is mainly a consequence of this lowering of temperature.

In dual fuel engines, the small quantity of pilot liquid fuel is injected into a mixture of gaseous fuel and air at mean temperature and pressure that may be different from the corresponding values for plain diesel operation. Hence, it would be expected that the processes of atomization, vaporization, and distribution of the small quantity of pilot fuel would be affected by any changes in the flow, thermal, and transport characteristics of the charge. Accordingly, there is a need to optimize appropriately the injection

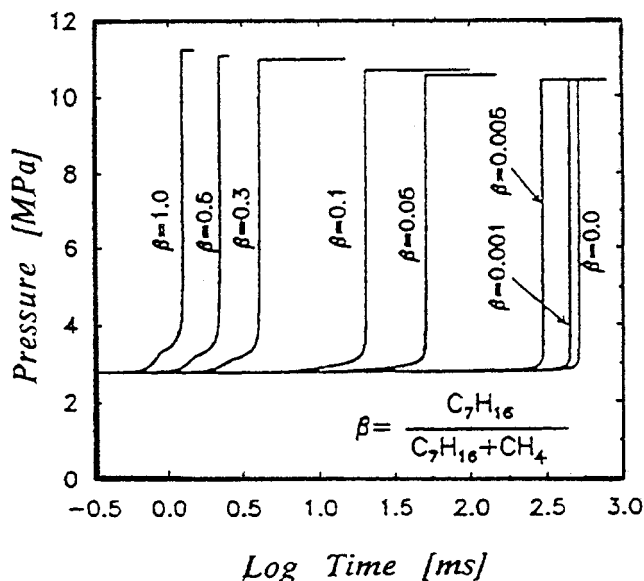


Fig. 11 Calculated temporal variations of pressure for different stoichiometric n-heptane methane mixtures with air for adiabatic constant volume reaction, $T_i = 800$ K and $P_i = 2.8$ MPa

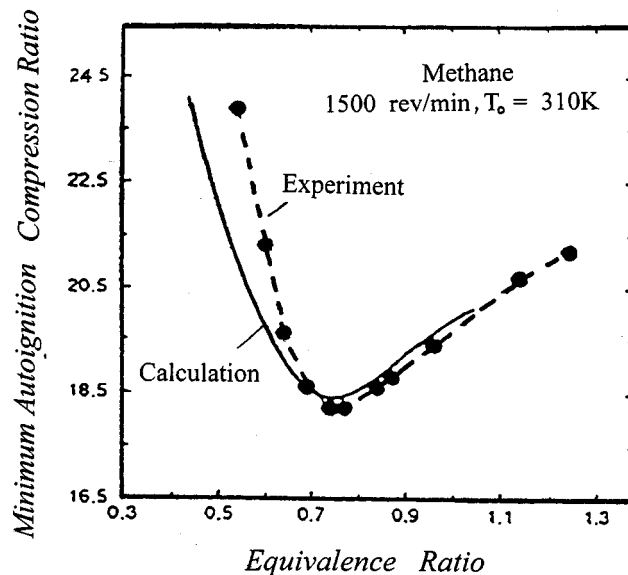


Fig. 12 Variations of the calculated minimum compression ratio for regular auto-ignition of different methane-air mixtures in a motored engine. The experimental data of Downs et al. [13] are shown.

characteristics for the dual fuel operation. Such a measure, often, is not followed because of the need to retain satisfactory full diesel operation whenever needed.

The pre-ignition reaction processes are widely different for the higher hydrocarbon pilot fuel vapor and for the gaseous fuel component such as methane. For example, n-heptane, a fuel which may represent adequately the behavior of the light fractions of diesel fuel, begins to undergo very complex multistage ignition reactions at relatively much lower temperature level than for the gaseous fuel methane which proceeds to ignition via simpler single stage. The presence of any liquid pilot fuel vapor in various regions of the gaseous fuel-air charge in a dual fuel engine, may bring about significant changes to the overall rates of pre-ignition reactions and the associated energy release rates, [11]. Figure 10 shows the very rapid reduction in the ignition delay and subsequent combustion time with increasing the mole fraction of heptane in its mixture with methane. These are calculated values us-

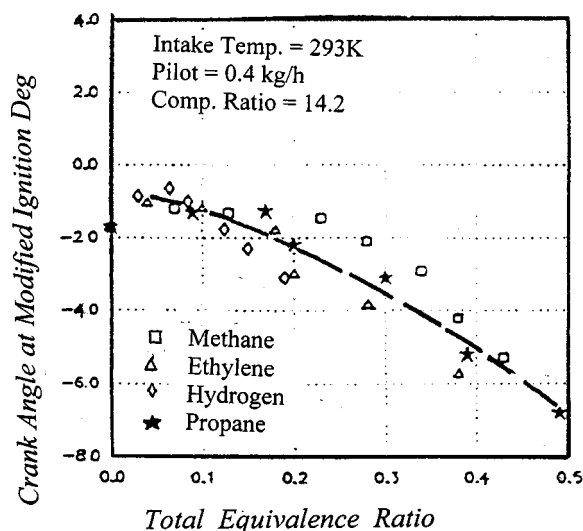


Fig. 13 Variations of the suitably modified ignition delay with total equivalence ratio for different fuels

ing fully detailed kinetics for constant volume conditions at initial temperatures and pressures that are comparable to those found near TDC in dual fuel engines. For the lower temperature case of 650 K, the reduction in these times is by a few orders of magnitude. Also, Fig. 11 shows that the presence of relatively minute quantities of heptane vapor with the methane does speed up the rates of ignition processes very substantially. The small concentration of the heptane begins to react with the oxygen well ahead of the methane producing both exothermic energy release and key transient products that serve to provide some of the radicals needed for the methane oxidation. Accordingly, the chemical interaction that goes on within the two fuel systems in various regions of the charge in a dual fuel engine must be sufficient to establish the progress of the combustion process, yet remains insufficiently known for the proper control of combustion and emissions.

A feature of the combustion process in dual fuel engines when operating on lean mixtures is that part of the gaseous fuel and some of the species produced in the combustion process can survive to the exhaust stage. These species surviving in very small concentrations through the residual gasses can play important chemical and thermal roles in the ignition and combustion processes of subsequent cycles, [12]. The thermal effect of the residual gases is reflected in changing the mixture temperature at the commencement of compression and influencing the temperature of the cylinder surfaces. The kinetic effect of these residuals may result in increasing the activity of preignition reactions during the latter stages of compression. The agreement between calculated and experimental values, [13], for the minimum compression ratio needed for autoignition of methane-air mixtures in the absence of pilot injection shown in Fig. 12, could be achieved only when full account of the thermal and kinetic effects of the residuals was made.

It can be seen that the length of the ignition delay in a dual fuel engine is affected by the presence of a gaseous fuel with the air during the ignition of the pilot liquid fuel. As shown in Fig. 13, when the effects of changes in the mean temperature of the charge and its partial pressure of oxygen due to the increased admission of the gaseous fuel, are corrected for, the delay period of the pilot follows a similar trend for the different fuels considered.

Combustion Under Light Load Conditions

The introduction of a gaseous fuel with the engine air, as shown earlier, can have a significant effect on the cylinder charge during compression affecting markedly the processes of pre-ignition and subsequent combustion of the pilot and the cylinder charge. The combustion zone enveloping a jet diffusion flame can extend significantly both radially and axially as the concentration of the gaseous fuel in the surrounding air is increased until a concentration is reached that would permit flame propagation to sweep through the entirety of the jet surroundings. This limiting concentration can be viewed as a simple function of the effective flammability limit of the gaseous fuel in air under the prevailing conditions. Moreover, it can be shown that the ignition of the droplets of the injected pilot will be affected similarly by the presence of the gaseous fuel and its concentration. Both the reaction time and the associated energy release will be modified increasingly by the raising of the concentration of the gaseous fuel in the droplet surroundings, [14].

Accordingly, an important general feature of the dual fuel engine is its relatively poor performance under light load condition when very lean mixtures of gaseous fuel in air are employed. The extent of this deterioration in performance depends largely on the pilot quantity employed, the gaseous fuel used, operating conditions and the engine employed. Under these conditions, as shown earlier, not only the ignition delay increases with the introduction of the gaseous fuel but also a significant proportion of the fuel will not burn completely, despite the presence of much excess air and a consistent pilot ignition. The flames originating from pilot igni-

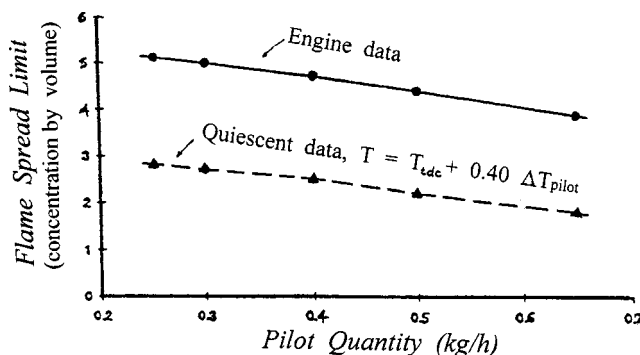


Fig. 14 Correlation of the observed lean operational limits in a dual fuel engine with the corresponding quiescent limits based on pilot combustion temperatures

tion regions cannot propagate in time far enough into the lean mixtures leaving some of the gaseous fuel unreacted to the exhaust stage. Normally, associated with this low gaseous fuel utilisation at lean mixtures is a significant increase in the carbon monoxide concentration well beyond that normally observed with the corresponding diesel operation, [15]. By using higher gaseous fuel concentrations the effective flammability limit will be reached when the fuel conversion becomes more complete and engine performance is substantially improved. As shown in Fig. 14, this mixture concentration limit can be correlated in terms of the quiescent flammability limit of the fuel in air at the mean temperature during pilot ignition. Thus, when converting diesel engines to dual fuel operation, the tendency is to retain diesel operation for idling and low load operations. Otherwise, the specific energy consumption then, (when based on the pilot diesel fuel and the gaseous fuel supplied) and the associated power output and emissions will be inferior to those of the straight diesel operation, Fig. 15. This is particularly evident for the relatively slow burning methane and at low intake temperature operation such as when using liquefied natural gas.

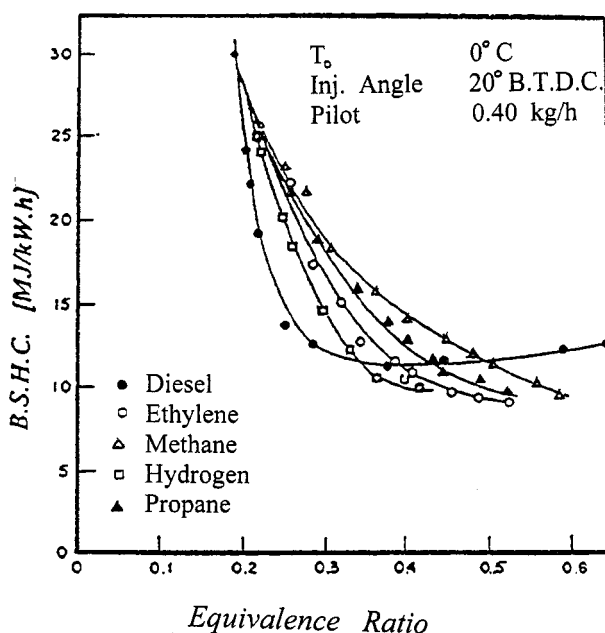


Fig. 15 Variation of the brake specific energy consumption (BSHC) for a dual fuel engine operating on different fuels at constant pilot quantity

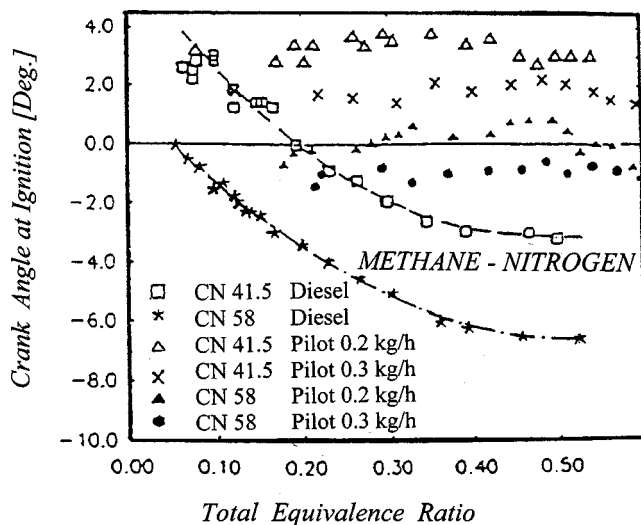


Fig. 16 Variations in the point of ignition with total equivalence ratio for different cetane number pilots with methane-nitrogen fuel mixtures

Normal dual fuel operation at light load when using relatively small pilot quantities can be improved by incorporating the following procedures, [16]:

- i. Using relatively large pilot quantities with optimum injection characteristics such as by employing the lowest injection nozzle opening pressure and advancing the injection timing somewhat without undermining diesel operation. As shown in Fig. 16, increasing the cetane number of the liquid pilot will also enhance light load operation and/or permit operation with gaseous fuel mixtures that may contain high concentrations of the diluents nitrogen or carbon dioxide, [9].
- ii. Partial restriction of the air component of the charge so as to produce an effectively richer mixture for the same mass of fuel gas. However, this throttling should be employed with care so as not to undermine pilot ignition. In turbocharged dual fuel engine applications at light load, excess air may be made to bypass the engine altogether. Also in multi-cylinder applications, "skip firing", where some cylinders may be operating wholly on the diesel mode while others operating as dual fuel but with less lean mixtures, may provide improved overall performance.
- iii. Slight preheating of the lean intake gas-air mixture or increasing the water jacket temperature provides a higher mixture temperature at the end of compression and reduces the fraction of the gaseous fuel surviving the combustion process. This is consistent with the tendency of the flammability limits of all common gaseous fuels to widen approximately linearly with the average mixture temperature. In turbo-charged engine applications, the feeding back of some warm excess air into the compressor intake at light load may also help.
- iv. The selective addition at light load of small amounts of a suitable auxiliary fuel to the main gaseous fuel methane supply, such as hydrogen, higher hydrocarbons or gasoline vapor, may improve dual fuel engine operation. Such an approach is to be avoided beyond the light load range since apart from affecting the specific energy consumption adversely, it can bring about an earlier onset of knock.
- v. Some stratification of the gaseous fuel component in relation to the air, if can be achieved effectively in practice, such as through direct gas injection, may improve light load operation through arranging for a slightly richer mixture in

the vicinity of the pilot fuel. Also, resort to some uncooled exhaust gas recirculation may also help. Optimum stratification also may reduce exhaust emissions very significantly.

Most probably, resorting to a judicious combination of these measures can be effective in improving the light load performance of the dual fuel engine while retaining the capacity to operate efficiently as a diesel, whenever needed.

Exhaust Emissions

The main constituents of the exhaust gas of dual fuel engines are unburned hydrocarbons in the form of unconverted gaseous fuel, carbon monoxide, carbon dioxide, oxides of nitrogen and particulate emissions. As was indicated previously, turbulent flame propagation from the ignition regions of the pilot will not proceed throughout the charge until the concentration of the gaseous fuel is beyond a limiting value that varies with operating conditions. When operating with a sufficiently lean charge a significant amount of the gaseous fuel and products of the pre ignition and partial combustion process survives to the exhaust stage. As shown, typically in Fig. 17, the concentration of the unconverted methane in the exhaust gas increases initially almost proportionally with the extent of gas admission. As the concentration of the gaseous charge approaches the limiting value the concentration of methane drops rapidly so that eventually it becomes well below the corresponding values usually encountered with gas fuelled spark ignition counterpart. Any unconsumed methane tends to be much less reactive than other hydrocarbons to produce photochemical smog. However, methane is known to be much more effective as a green house gas than carbon dioxide and will require control. As shown in Fig. 18, the relative size of the pilot employed as well as its injection characteristics has a controlling influence on the emission of methane at any equivalence ratio. It can be seen that relatively large pilots need to be used at very light load but they may be reduced very substantially in size at high load. The recent increase in the practice of using direct injection of the gaseous fuels in multi-cylinder installations and for two stroke engines, though increases complexity can permit a further substantial reduction in the exhaust emissions of methane.

As part of the sequencing of the oxidation reactions of methane in air, carbon monoxide is produced in the early stages of the oxidation. Much of this carbon monoxide oxidises further on with time to carbon dioxide. For very lean mixtures some of the carbon monoxide may not be oxidized further due to the failure of the combustion process to encompass the whole charge and the relative slowness of the reaction processes. Hence, at very light load,

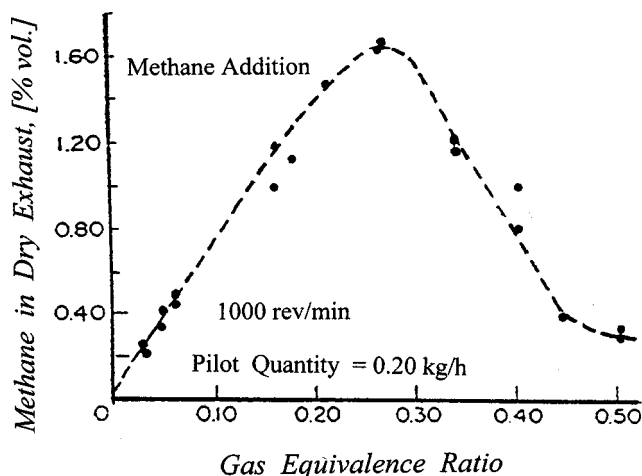


Fig. 17 Variations of the exhaust concentration of unreacted methane with the intake gas equivalence ratio at a constant pilot quantity with methane operation

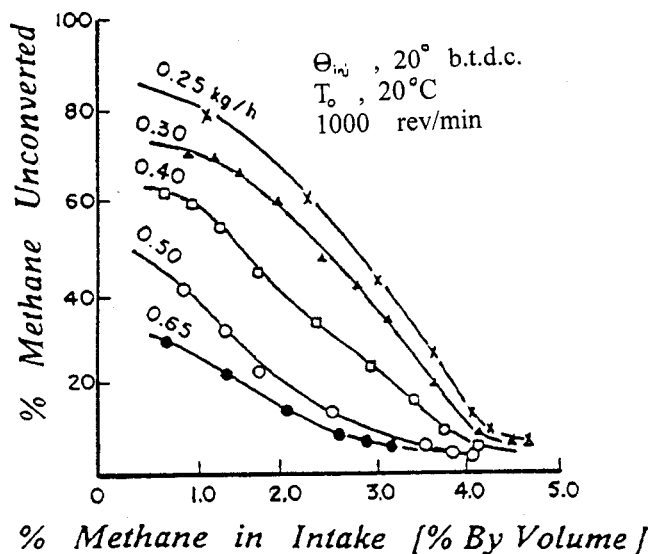


Fig. 18 Variation of the unconverted methane in the exhaust gas of a dual fuel engine with methane addition for different pilot quantities

when very lean mixtures and especially when small pilot quantities are involved, some carbon monoxide may survive to the final exhaust phase, as shown typically in Fig. 19. The percentage of the total carbon present in both the pilot fuel and the methane that eventually appears as carbon monoxide in the exhaust gas remains relatively small comparable to the concentrations normally found in diesel operation, [15].

The carbon monoxide in the exhaust gases of dual fuel engines originates largely from the gaseous fuel component in regions mainly within and adjoining the burning pilot. Thus, the concentration of carbon monoxide will be dependent on the size of these regions. As shown in Fig. 20, for the same equivalence ratio, the use of a larger pilot tends to produce higher carbon monoxide concentrations. The increase in intake charge temperature as shown in Fig. 21, will enhance the oxidation of carbon monoxide within the time available. Changes in the injection characteristics of the pilot for the same pilot mass and equivalence ratio can modify the extent of methane conversion and the exhaust concentrations of carbon monoxide. The contribution of the pre-ignition

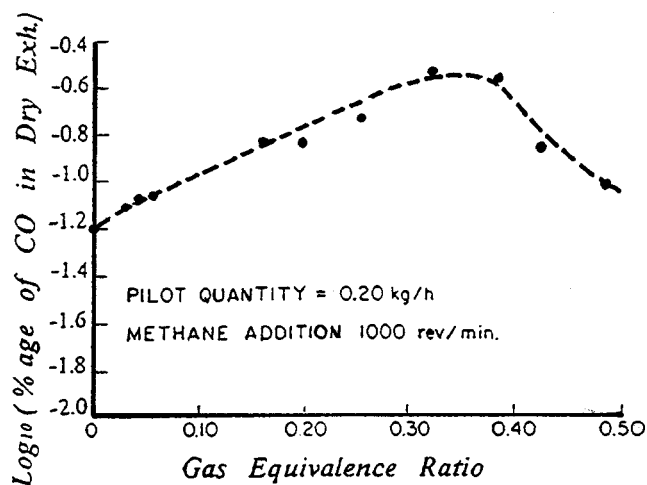


Fig. 19 Variation in the exhaust concentration of carbon monoxide with changes in the gas equivalence ratio for a constant pilot quantity with methane operation

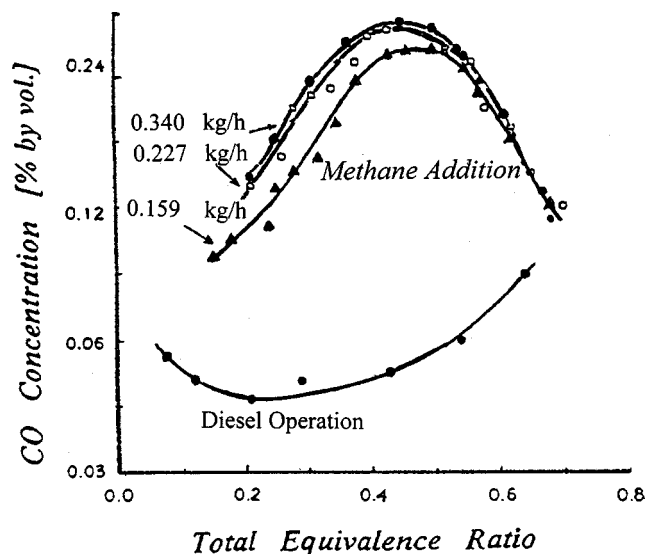


Fig. 20 Logarithmic variations of the exhaust concentration of carbon monoxide with the total equivalence ratio for different pilot quantities with methane operation

reactions of the bulk of the charge to carbon monoxide exhaust emissions, tends to be variable but remains generally small, [17].

The production of oxides of nitrogen depends primarily on the peak value of the combustion temperature, the effective volume of the combustion zone, the availability of oxygen and on whether sufficient time for the oxygen-nitrogen reactions is available to proceed to significant levels of completion. In dual fuel engines, much of the NO_x production is associated with the pilot zone where very high local temperatures are achieved and longer reaction times are possible. Some further NO_x production, but to a much lesser extent will be also from heated mixture regions in the vicinity of the pilot combustion zone. Increasing the pilot will have an important contribution to the increased production of NO_x as the size of the combustion zone is similarly increased. With very lean mixture operation, relatively very little NO_x production is expected from regions of the rest of the gaseous mixture charge. As shown typically in Fig. 22, the production of oxides of nitro-

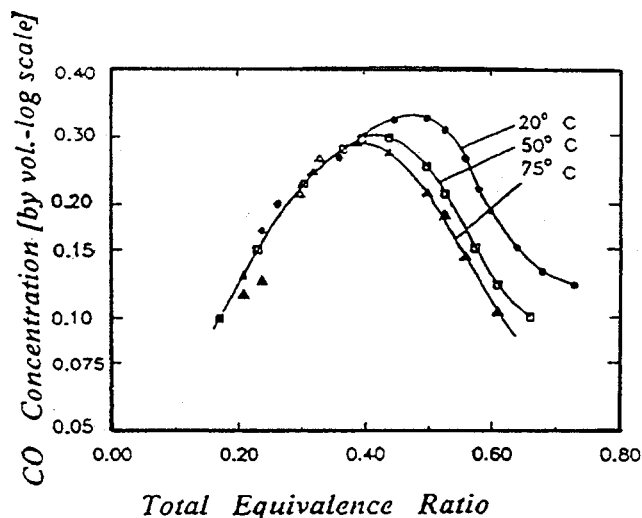


Fig. 21 Variation of the volumetric exhaust concentration (log scale) of CO with total equivalence ratio at a constant pilot quantity with methane operation for three different intake mixture temperatures

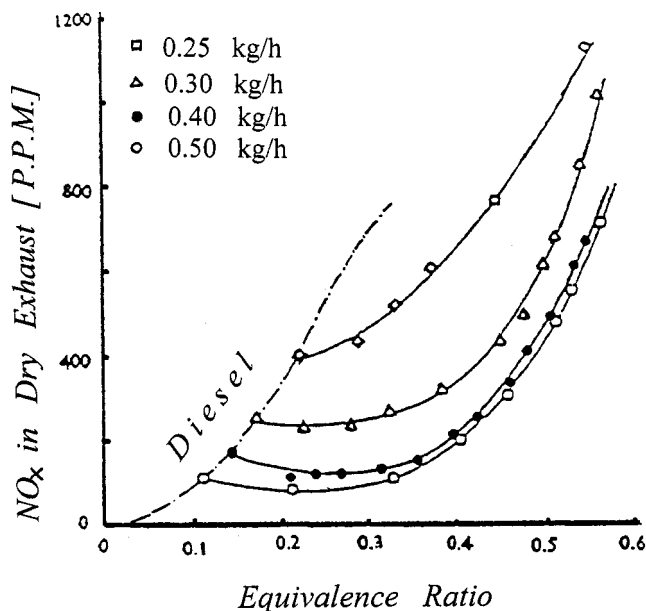


Fig. 22 Variations in the exhaust concentration of oxides of nitrogen with total equivalence ratio for different pilot quantities with methane operation

gen is influenced markedly by both the quantity of the pilot employed and the overall equivalence ratio. Resorting to operating dual fuel engines on relatively large pilots at light load with very lean gaseous fuel-air mixtures and as the load is increased on increasingly smaller pilots, will maintain the oxides of nitrogen at levels below those encountered in the corresponding diesel operation. With small pilots and the overall lean mixture operation of dual fuel engines the level of oxides of nitrogen produced will be also lower than in comparable spark ignition operation, so long as the onset of knock is avoided.

Since the bulk of the energy release in dual fuel operation comes from the combustion of the gaseous fuel a very important feature is that the extent of smoke and particulate emissions will be very much lower than those encountered with the corresponding diesel operation. This is achievable over the whole power output range, even when very low intake mixture temperatures are involved, such as with LNG operation. This feature makes the conversion of diesel engines to dual fuel engine operation distinctly attractive over the whole load range.

Knock in Dual Fuel

The energy release in dual fuel engines normally involves the combined contribution of the pilot fuel regions together with the rapid turbulent flame fronts propagation originating from these regions. This mode of combustion that combines autoignition and rapid flame propagation processes is primarily responsible for the ability of dual fuel engines to burn mixtures that are much leaner than those normally possible in gas fuelled spark ignition engines. A prime requirement of any alternative gaseous fuel for satisfactory operation in dual fuel engines is that its mixture with air would not autoignite spontaneously during or following the rapid pilot energy release. Failing to do so can lead to the onset of knock which manifests itself with excessively rapid rates of pressure rise and overheating of the walls leading to significant loss in efficiency with increased cyclic variations. Persistent knock is highly objectionable and needs to be avoided; otherwise it may lead to engine failure.

Knock in dual fuel operation is of an autoignition nature involving mostly the gaseous fuel-air mixture. Depending on the size of the pilot and its mode of injection, knock can be perceived to involve auto-ignition of that portion of the charge in the neighbor-

hood of these ignition centres leading to very high rates of pressure rise and a consequent very rapid burning of the remaining parts of the charge. With much smaller pilots, the energy release during the initial stages of ignition and the resulting turbulent flame propagation can lead under certain conditions to an auto-ignition of the charge well away from these ignition centres in end gas regions ahead of these flames, in a manner that resembles the occurrence of knock in spark ignition engines.

The knock limited power output for any fuel and pilot setting has been shown to deteriorate logarithmically with the inverse of the intake absolute temperature. Accordingly the onset of knock may be avoided or curtailed through a number of design and operational measures such as lowering the induction and water jacket temperatures and delaying the commencement of pilot injection. Other effective measures such as lowering the compression ratio are usually not resorted to since they would undermine the operation of the engine as a diesel.

One of the prime means of avoiding the onset of knock in a natural gas fuelled dual fuel engine is through careful control of the composition of the gaseous fuel mixture. This can be accomplished through measures such as reducing to a minimum the concentration of any higher hydrocarbon components with the methane and/or through increasing the concentration of any diluents such as carbon dioxide or nitrogen that may be present with the fuel mixture and avoiding having droplets of fuel condensates with the original gas. Of course, in principle an effective procedure for reducing the incidence of knock is through optimum distribution of the gaseous fuel to air ratio by a proper stratification of the gaseous fuel component within the cylinder such as via timed gaseous fuel injection during the latter parts of the compression stroke or to a lesser degree of effectiveness via its timed injection just outside the cylinder. This stratification aims at having less reactive leaner gaseous fuel-air mixtures in regions away from the ignition centres of the pilot where richer mixtures are located. This would require careful control of the gaseous fuel distribution such that it will not hinder pilot fuel ignition nor encourage the autoignition of regions adjacent to the pilot while reducing the tendency of the charge away from these centres to autoignite despite the high rates of increase in cylinder pressure and temperature, [18]. Fortunately, the knocking operational regions with methane are normally beyond most common operational conditions unless highly supercharged large bore engines and unoptimized pilot quantity and injection characteristics are employed. However, for some other gaseous fuel mixtures containing the more reactive higher hydrocarbons, the occurrence of knock in unsuitably modified diesel engines is often encountered. Enhancing the cetane number of the pilot fuel can have a beneficial effect by delaying a little the onset of knock.

A relatively simple approximate approach for the adequate prediction of the onset of knock for any dual fuel engine set up assumes knock to be associated with autoignition of the gaseous fuel-air charge at a mean temperature and pressure that correspond to the end of rapid combustion of the pilot at constant volume while accounting fully for the preignition reaction activity of the charge. Further refinements of such an approach consider more realistically other important operational variables such as a more progressive burning of the pilot, residual gas effects and heat transfer. Further refinements to knock modelling is an area of intense research activity, especially since there is always a need not only to avoid the onset of knock but also to do so while maintaining high outputs and efficiencies with acceptable levels of exhaust emissions.

Concluding Remarks

The dual fuel engine, not only will consume a wide range of gaseous fuel resources effectively, but also has the potential to avoid much of the current and future problems facing the diesel engine including the need for very significant reductions in its exhaust emissions. There is a continuing need to improve our

understanding of the combustion processes so as to carry out the conversion of a wide range of diesel engines to operate efficiently, cleanly and economically on a wide range of gaseous fuels. Obviously, for any engine and fuel systems, a multitude of variables needs to be controlled simultaneously and optimally to enhance light load performance, reduce ignition delays, avoid knock and reduce exhaust emissions over the whole load range. There is always a need to evaluate the consequences of a number of control approaches, adopt the positive features that may be identified while keeping any increase in engine complexity and cost to low levels. This is being achieved through employing a combination of effective computational procedures aided and supported by experimental testing.

Acknowledgment

The author has drawn very freely in this contribution from the research work of his various past and current associates. He acknowledges gratefully such association and their generous contributions.

References

- [1] Turner, S. H., and Wearer, C. S., 1994, "Dual Fuel Natural Gas/Diesel Engines," Gas Research Institute, No. GRI-94/0094.
- [2] Karim, G. A., 1987, "The Dual Fuel Engine" in Robert L. Evans (Ed) "Automotive Engine Alternatives," Plenum Press, N.Y.
- [3] Karim, G. A., and Amoozegar, N., 1983, "Determination of the Performance of a Dual Fuel Diesel Engine With the Addition of Various Liquid Fuels to the Intake Charge," S.A.E. Paper No. 830265.
- [4] Danyluk, P. R., 1993, "Development of a High Output Dual Fuel Engine," ASME Paper No: 93-ICE-20.
- [5] Hill, P. G., and Douville, B., 1998, "Analysis of Combustion in Diesel Engines Fuelled by Directly Injected Natural Gas," ASME Paper No. ICE 30-3.
- [6] Wong, Y., and Karim, G. A., 2000, "A Kinetic Examination of the Effects of Recycled Exhaust Gases on the Autoignition of Homogeneous N-Heptane-Air Mixtures in Engines," S.A.E. No. F1236.
- [7] Liu, Z., and Karim, G. A., 1998, "An Examination of the Ignition Delay Period in Gas Fueled Diesel Engines," ASME J. Eng. Gas Turbines Power, **120**, pp. 225–231.
- [8] Badr, O., Karim, G. A., and Liu, B., 1998, "An Examination of the Flame Spread Limits in a Dual Fuel Engine," Appl. Therm. Eng., **19**, pp. 1071–1080.
- [9] Karim, G. A., Kibriya, M. G., Lapucha, R., and Wierzbza, I., 1988, "Examination of the Combustion of a Fuel Jet in a Homogeneously Pre-Mixed Lean Fuel-Air Stream," S.A.E. Paper No. 881662.
- [10] Wierzbza, P., Karim, G. A., and Wierzbza, I., 1992, "An Analytical Examination of the Combustion of a Turbulent Fuel in an Environment Containing Pre-mixed Fuel or a Diluent and Air," ASME J. Energy Resour. Technol., **117**, pp. 234–239.
- [11] Khalil, E. B., and Karim, G. A., 2002, "A Kinetic Investigation of the Role of Changes in Composition of Natural Gas in Engine Applications," ASME J. Eng. Gas Turbines and Power, **124**, pp. 404–411.
- [12] Liu, Z., and Karim, G. A., 1996, "An Examination of the Role of Residual Gases in the Combustion Processes of Motored Engines Fuelled With Gaseous Fuels," S.A.E. Paper No. 961081.
- [13] Downs, D., Walsh, A. D., and Wheeler, R. W., 1951, "A Study of the Reactions That Lead to Knock in the Spark Ignition Engine," Philos. Trans. R. Soc. London, Ser. A, **243**, pp. 463–524.
- [14] Samuel, P., and Karim, G. A., 1994, "An Analysis of Fuel Droplets Ignition and Combustion Within Homogeneous Mixtures of Fuel and Air," SAE Paper No. 940901.
- [15] Karim, G. A., Liu, Z., and Jones, W., 1993, "Exhaust Emissions From Dual Fuel Engines at Light Load," S.A.E. Paper No. 932822.
- [16] Karim, G. A., 1991, "An Examination of Some Measures for Improving the Performance of Gas Fuelled Diesel Engines at Light Load," S.A.E. Paper No. 912366, S.A.E. Transactions 1991.
- [17] Liu, Z., and Karim, G. A., 1997, "Simulation of the Combustion Processes in Gas Fuelled Diesel Engines," Proceedings—Inst. of Mech. Engineers, J. of Power and Energy, **211**, pp. 159–171.
- [18] Karim, G. A., and Lui, Z., 1992, "A Prediction Model of Knock in Dual Fuel Engines," S.A.E. Paper No. 921550, S.A.E. Transactions 1992.
- [19] Gunee, C., Razavi, M. R., and Karim, G. A., 1998, "The Effects of Pilot Fuel Quality on Dual Fuel Engine Ignition Delay," SAE Paper No. 982453.
- [20] Khalil, E., Samuel, P., and Karim, G. A., 1996, "An Analytical Examination of the Chemical Kinetics of the Combustion of N-Heptane-Methane Air Mixtures," S.A.E. Paper No. 961932.

J. Hiltner

Hiltner Combustion Systems,
108 Fairbanks Road, Unit #1,
Farmington, ME 04938
e-mail: jdh@hiltnercombustionsystems.com

R. Agama

Caterpillar, Inc.,
Mossville Engine Center,
Mossville, IL 61552-0600
e-mail: agama_rey@cat.com

F. Mauss

Division of Combustion Physics,
Lund Institute of Technology,
22100 Lund, Sweden
e-mail: Fabian.mauss@forbrif.lth.se

B. Johansson

e-mail: bengt.johansson@vok.lth.se

M. Christensen

e-mail: magnus.christensen@vok.lth.se

Department of Heat and Power Engineering,
Lund Institute of Technology,
22100 Lund, Sweden

Homogeneous Charge Compression Ignition Operation With Natural Gas: Fuel Composition Implications

Homogeneous charge compression ignition (HCCI) is a potentially attractive operating mode for stationary natural gas engines. Increasing demand for efficient, clean burning engines for electrical power generation provides an opportunity to utilize HCCI combustion if several inherent difficulties can be overcome. Fuel composition, particularly the higher hydrocarbon content (ethane, propane, and butane) of the fuel is of primary concern. Fuel composition influences HCCI operation both in terms of design, via compression ratio and initial charge temperature, and in terms of engine control. It has been demonstrated that greater concentrations of higher hydrocarbons tend to lower the ignition temperature of the mixture significantly. The purpose of this paper is to demonstrate, through simulation, the effect of fuel composition on combustion in HCCI engines. Engine performance over a range of fuels from pure methane to more typical natural gas blends is investigated. This includes both the impact of various fuels and the sensitivity of engine operation for any given fuel. Results are presented at a fixed equivalence ratio, compression ratio, and engine speed to isolate the effect of fuel composition. Conclusions are drawn as to how the difficulties arising from gas composition variations may affect the future marketability of these engines. [DOI: 10.1115/1.1581895]

Introduction

Homogeneous charge compression ignition (HCCI) describes an engine operating regime in which compression heating of the charge results in ignition throughout the bulk of a largely homogeneous mixture. High rates of dilution, via either lean equivalence ratio or high EGR rates, provides both excellent indicated thermal efficiency and low specific NO_x emissions. The primary impediments that have limited the market penetration of these engines are: low power density, difficulty in control, limited load/speed range, and high emissions of unburned hydrocarbons and carbon monoxide.

Homogeneous charge compression ignition (HCCI) engines were relatively common in the early part of the last century (1900–1920) in the form of “oil vaporizing engines” designed to burn heavy liquid fuels that were not easily premixed with air prior to ingestion, [1]. With the arrival of improved diesel fuel injection equipment, these low efficiency [compression ratio ~ 3.5] and somewhat unmanageable engines passed into obscurity. The concept of compression ignition of a well mixed charge continued in the form of model engines designed to run at very high speed. These engines typically utilize a glow plug or adjustable compression ratio for starting, but run without an ignition source when warm. In both these cases, it was neither high efficiency nor low emissions which called for HCCI operation, but rather simplicity of design and the ability to produce fast, repeatable combustion.

Beginning in the late 1960s, researchers from three disparate areas of the engine community began, somewhat by accident, investigating HCCI combustion. A number of diesel engine manu-

facturers were, at this time, investigating the use of a wide range of fuels in diesel engines. When low cetane fuels were applied to traditional diesel engines, high levels of premixing occurred prior to ignition, resulting in the first branch of modern HCCI engines. This led eventually to the MK combustion system produced by Nissan, [2]. Around the same time, chemical kinetic and experimental investigations of engine knock were revealing that under some highly diluted operating conditions, an engine would run smoothly, efficiently and with low emissions if the proper combination of compression ratio, fuel, and intake conditions were maintained. This path led to the current stream of researchers investigating port-injected HCCI engines utilizing high octane number fuels ([3–7], for example). The third research track was established by Japanese engineers investigating the cause of run-on in two stroke engines. These efforts determined that autoignition of the charge due to high levels of elevated temperature residuals was responsible for the bulk ignition. This realization led to the Honda ATAC system in current production, [8].

This paper is concerned with the application of port-injected four-stroke HCCI technology to stationary natural gas engines. Open and pre-chamber engines in production are capable of a thermal efficiency from 35% to 42%, with BMEP levels from 14 to 18 bar, and NO_x emissions from 75 to 150 PPM. Homogeneous charge compression ignition technology shows excellent potential for penetration into this market for several reasons. The combination of high thermal efficiency with very low NO_x emissions could allow current reciprocating engines to meet future emissions regulation without expensive NO_x aftertreatment. Engines in this market typically operate over only a very narrow range of load and speed, greatly simplifying the control issues associated with HCCI. Finally, natural gas has been shown through both simulation and experiment to be well suited to HCCI applications.

A number of research institutions are currently working with natural gas HCCI engines through both simulation, [9–15], and

Contributed by the Internal Combustion Engine Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the ICE Division May 2001; final revision received by the ASME Headquarters October 2001. Associate Editor: D. Assanis.

experiment, [3–7,16]. These investigations have demonstrated that NO_x emissions of 5–10 ppm can be realized at indicated thermal efficiencies approaching 50% and IMEP values of up to 14 bar. The primary production limitations of these engines to date are the low achievable IMEP for a given peak cylinder pressure, high emissions of unburned hydrocarbons and carbon monoxide (which are difficult to oxidize at the given exhaust temperatures) and concerns about engine control.

This paper addresses a number of these concerns by means of a single-zone chemical kinetics based cycle simulation. The temptation that exists with engine simulation is to exercise the simulation simply to prove that it is capable of reproducing engine test cell results. While this is an important first step and can lead to significant cost savings, it is not the primary purpose of modeling work. This paper focuses on using simulation methods to better understand current test results and to investigate operating conditions which may be difficult to achieve in current research engines.

The chemical composition of natural gas has pronounced effects on combustion in HCCI engines, [7]. This impacts not only the required inlet conditions and compression ratio but also complicates issues related to engine control. The role of individual components is investigated in a similar way to previous efforts, but further work is directed towards determining the sensitivity of the engine to inlet manifold temperature for any given mixture. It is shown, through simulation, that potential exists to improve the power density of HCCI engines by locating an optimum tradeoff between thermal efficiency and power ratio. The term power ratio here refers to the ratio between the full cycle indicated mean effective pressure and the peak cylinder pressure. This ratio provides a metric of work that can be produced for a given peak cylinder pressure. Finally, extensive discussions within the engine community continue to focus around the existence of “active radicals” and their role in HCCI combustion. While radicals themselves are unable to survive the exhaust, intake, and compression strokes to influence the next combustion event, this paper demonstrates that partially oxidized fuel from misfiring cycles can have a definite impact on ignition of the charge, having a profound effect on engine stability under some operating conditions.

Simulation Method and Conditions

In order to investigate the behavior of natural gas HCCI engines, a single-cylinder cycle simulation based on chemical kinetics has been developed. The model assumes that the cylinder contents may be represented by a single temperature and chemical composition throughout the mixture. This single-zone type of simulation has been described extensively in the literature, [9,15]. The reader is referred to these publications for both the derivation of the model and an extensive discussion of its shortcomings. Needless to say, the cylinder contents are neither perfectly mixed, nor of uniform temperature. Nonetheless, single-zone simulations provide an excellent match with experimental results in terms of overall engine performance trends and particularly start of combustion prediction. The primary shortcomings are an inability to accurately predict hydrocarbon and carbon monoxide emissions and under prediction of the heat release duration by up to 50%. Single zone models also tend to overpredict NO_x due to the under prediction of heat release duration and the exponential dependence of NO_x formation on temperature. Despite these weaknesses, the speed and robust nature of the single-zone model make it an ideal tool for analyzing engine operating regimes.

The model utilized here is based on a chemical kinetics code developed at the Lund Institute of Technology by one of the authors. Details of the model and its validation with experimental engine data is provided elsewhere, [9]. Several additions have been made to the model for this work. While previously the model covered only the valve closed period, the extended simulation contains filling and emptying models that can accurately capture the gas exchange processes and predict the residual gas fraction

Table 1 Engine configuration and simulation parameters

Engine Geometry	
Bore (mm)	120.65
Stroke (mm)	140
Connecting rod length (mm)	260
Compression ratio	19.8:1
Intake valve diameter (mm)	50
Intake valve lift (mm)	11.9
Exhaust valve diameter (mm)	46
Exhaust valve lift (mm)	13.4
Simulated Operating Conditions	
Engine speed (rev/min)	1000
Cylinder wall temperature (Kelvin)	390
Intake manifold pressure (bar absolute)	1 and 3
Exhaust Manifold pressure (bar absolute)	1 and 3
Volume fraction methane in fuel (%)	85 to 100
Volume fraction ethane in fuel (%)	0 to 15
Volume fraction propane in fuel (%)	0 to 15
Volume fraction butane in fuel (%)	0 to 15
Intake manifold temperature (Kelvin)	350 to 450
Simulation Parameters	
Simulation type	single-zone
Number of species considered	55
Number of reactions	596
Solution method	implicit
Maximum step size (CAD)	0.5
Typical steps per cycle	2000
Computation time per cycle (seconds)	15
Typical cycles to convergence	5

and composition. As shown later, the residual gas acts as a source of thermal, and sometimes chemical energy. Some basic parameters of both the engine geometry and simulation structure are given in Table 1.

The geometry described above is based on a single cylinder engine in use in the engine laboratory at Lund. All simulations are run at a constant speed of 1000 rpm, with varying inlet conditions and gas compositions. The chemical kinetic scheme is based on that described in previous publications, with the addition of the extended Zeldovich mechanism to provide a qualitative measure of NO_x emissions. It is difficult, in a single-zone model, to accurately capture the NO_x formation as the heat release tends to be more rapid than seen in real engines, resulting in higher peak temperatures and thus higher NO_x formation rates.

An implicit scheme devised by one of the authors provides for highly efficient solution of the set of 57 coupled differential equations (55 species concentrations, cylinder gas temperature, and total cylinder gas mass). Under typical running conditions, the simulation requires four to five cycles to reach convergence. Convergence is defined as two consecutive cycles whose peak temperature differs by less than 1 K. Heat transfer within the model is handled by a simple Woschni relation type function. The model allows a full range of natural gas compositions consisting of any combination of methane, ethane, propane, and butane. The simulation can be run for batches of different conditions and is capable of finding the optimum operating point given a set of initial constraints. As the potential combinations of equivalence ratio, fuel composition, intake pressure, intake temperature, and engine geometry are endless, a narrow range of operating conditions at a fixed engine geometry have been selected to demonstrate the fundamental performance of HCCI engines utilizing natural gas of varying composition. While this narrow range of operating condition allows the examination of various trends (as discussed below), extrapolation of these results to other applications, geometries, and operating conditions should be undertaken with caution.

Two-Component Fuel Behavior

While typical natural gas is a mixture of methane and a variety of other higher hydrocarbons, a logical first step in attaining an understanding of its behavior in HCCI engines is to simulate engine performance with simple mixtures of methane and one addi-

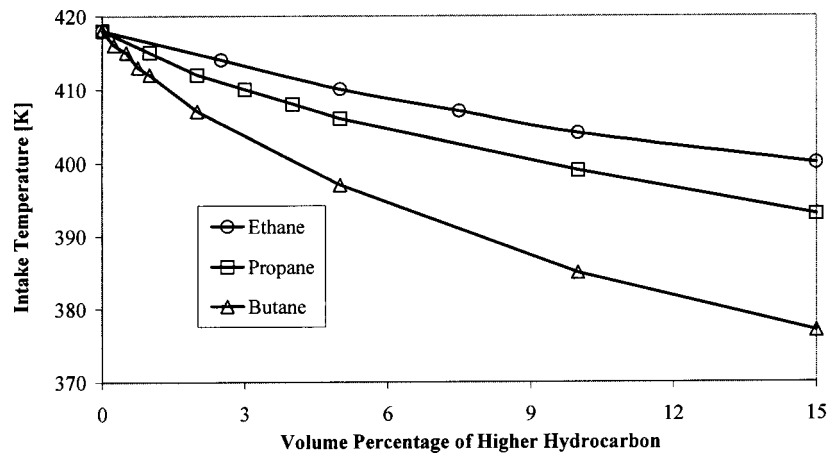


Fig. 1 Intake temperature for peak efficiency of two-component fuels

tional component. Similar work has been presented recently, [11], and has demonstrated that each additional component tends to lower the required intake temperature for peak thermal efficiency, with ethane providing the least and butane the most reduction.

The current simulation has been run at a set operating condition with a range of two-component fuels to demonstrate the functionality of the code and agreement with previous work. For the sake of simplicity, a naturally aspirated operating condition (intake pressure = 1 bar) was chosen, along with a fixed equivalence ratio ($\phi = 0.3$). While the dependence of performance on fuel composition shown in this section is directly applicable only to this condition, the operating trends are consistent with laboratory experience by the authors and others in this field.

A range of mixtures have been tested from pure methane at one extreme to 15% by volume addition of each of the other fuels considered (ethane, propane, butane). In each case, the simulation was run in batch mode where the intake manifold temperature was varied until the peak thermal efficiency with the given fuel was found. The resolution used for this determination was 1 K, in order to shorten simulation time.

Figure 1 gives the results of this initial study, with the intake temperature required for peak thermal efficiency plotted as a function of the volume percentage of the additive fuel. This graph captures nearly the full range of commercially available natural gas ranging from LNG at the very pure end to low methane number gases at the other extreme. Immediately evident from the graph is the varying impact of the gases, with the higher molecular weight butane showing a much more pronounced affect than ethane or propane. This behavior is interesting in the context of typical gas compositions which are typically much higher in ethane and propane volume fraction than in butane. It does indicate the even the relatively small concentrations of butane found in most gas reserves (1–2% by volume) will be significant in terms of HCCI combustion.

The primary implication of Fig. 1 is that the design of an HCCI engine, even for stationary use, is likely to require explicit development effort for any predetermined gas quality. This will likely take the form of fixing the compression ratio so as to achieve a somewhat constant intake manifold temperature across a range of engines designed for different gas compositions. The other possibility being a fixed compression ratio engine where the inlet manifold temperature controls (EGR, heat exchange with the exhaust etc.) are adapted to the specific application. In any case, some concerted design effort will be required for each application and the engine in the field will likely be significantly more sensitive to gas quality than a traditional spark ignited engine is to methane number. The result may be additional “de-rating” of the engine as is done with detonation margin in current engines.

Engine Fuel Sensitivity

While Fig. 1 reveals a good deal about engine performance with different two component gas compositions, it does not specify how the engine performs as the intake temperature deviates from the temperature ideal in terms of thermal efficiency at a given load and speed. This may be termed the sensitivity of the engine. A robust spark-ignited engine will operate over a wide range of intake manifold temperatures with almost no lower limit (typically a mechanical limit for gaseous fuel engines) and the upper limit determined by detonation. Thus a modern stationary gas engine would have a specified upper ambient temperature limit, above which engine performance will be compromised due to ignition adjustments required to avoid knock.

The situation is much more severe in an HCCI environment. The lower limit of operation is a relatively sharp drop in mass fraction burned, basically attributable to a failure of the “bulk ignition” event. While there is not an upper “knock limit” (assuming that the charge is suitably diluted) the peak cylinder pressure and NO_x emissions both increase dramatically as the intake manifold temperature is increased.

This behavior is demonstrated in Fig. 2. The base conditions are once again an equivalence ratio of 0.3, an intake pressure of one bar and an engine speed of 1000 rpm. Four different fuels have been simulated, ranging from pure methane at one end of the ignition spectrum to 85% methane and 15% butane (by volume) at the other extreme. For each fuel, the simulation has been run over a wide range of intake manifold temperature. The general trend for each fuel is quite clear. Beginning with at low intake temperature, one can follow the figure to the right to see the effect of increasing intake manifold temperature. Below some base temperature there is minimal heat release and the mixture is essentially inert. This leads to very low or negative thermal efficiency (i.e., the work done by the partially combusted gases is insufficient to overcome pumping work) and also very low or negative power ratios. As the intake temperature is increased from this point some partial oxidation of the charge begins. This upward sloping of the indicated efficiency is much more pronounced in the butane laden mixtures as would be expected. It is possible in these cases to release up to 20% of the energy of the fuel without causing bulk ignition of the primary fuel (methane). In the pure methane case, this effect is much less pronounced, and only a small fraction of the heat may be released without causing overall bulk ignition.

If the intake temperature increases further, a region of unstable operation is encountered for all fuels. This range is shown as a break in each thermal efficiency graph covering a range of five to seven degrees in inlet manifold temperature. This range of insta-

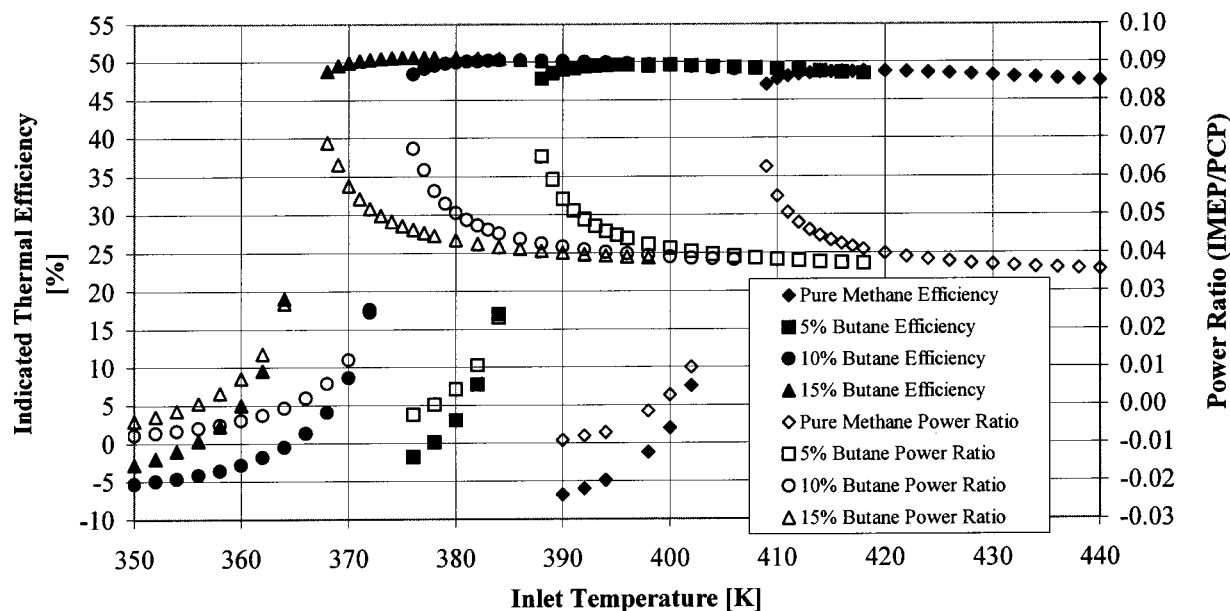


Fig. 2 Operation with methane/butane mixtures

bility tends to be more pronounced the higher the methane content of the fuel. Operation in this region is discussed in detail in a later section of the paper.

When the intake manifold temperature reaches some critical value, ignition of the main charge occurs. At the lowest of these temperatures, ignition occurs well after TDC. This results in some loss of thermal efficiency compared to the peak value due to a reduction in the degree of constant volume combustion. Despite this, there are many advantages to operating on this very ragged edge of stability.

The primary advantage is in terms of the “power ratio” plotted on the secondary y-axis. This parameter, here specified as the ratio between the full cycle IMEP (IMEP including the intake and exhaust strokes as well as compression and expansion) and the peak cylinder pressure, is a measure of how efficiently the engine is using the available cylinder pressure range. For example, a typical spark ignited natural gas engine with an allowable peak cylinder pressure of 180 bar and an IMEP of 14 bar provides a power ratio of just under 0.08. Typical diesel engine values would be closer to 0.10. It can be seen from the graph that if the engine is operated near the stability limit, values above 0.07 can be obtained. On the other hand, increasing the intake manifold temperature by a mere 5 K moves this value well below 0.05 and outside of the production feasible range. While this small increase in intake temperature greatly decreases the achievable power density of the engine, it has little effect on the indicated thermal efficiency. The peak thermal efficiency occurs several degrees above the terminal temperature as the heat release event is moved closer to TDC and also becomes more rapid. Thus, while the engine can be operated over a range of intake temperatures, the ideal operating range is, in fact, very narrow. The power ratio figures given here are lower than would be seen in experiments due to the over prediction of heat release rate by the single zone model.

Figure 3 is provided to clarify the cause of the increased power ratio that is obtainable near the stability limit of the engine. Shown in the figure are cylinder pressure traces obtained for the 10% butane mixture from Fig. 2. At 350 K there is little chemical activity, providing what is essentially a motored trace. Notice that this trace lies slightly above the others during the compression stroke due to the reduced heat transfer resulting from reduced inlet

temperature. This behavior is important in HCCI engines, where high intake temperatures produce low volumetric efficiency and thus low power density in unboosted operation.

The next pressure trace represents the lowest stable operating temperature (376 K inlet temperature). The reason for the increased power ratio (from $T = 382$ K to $T = 376$ K) is clear, as the heat release is clearly occurring well after top dead center and thus the peak cylinder pressure is dramatically reduced in the later case. The late timing of the ignition event does lead to some deterioration of thermal efficiency, but this effect is relatively minor as the piston motion around TDC is so slow. Experimental results reveal that late combustion always has a higher brake efficiency than early combustion, due to heat transfer and other secondary effects. Three additional cylinder pressure traces are shown with progressively increased intake manifold temperature. The primary point here being that a small increase in intake manifold temperature (from 376 to 382 K) increases the peak cylinder pressure by nearly 40%.

There are a number of factors which prevent operation in this ideal intake temperature range. First, this model, being single zone, does not capture wall effects that increase unburned hydrocarbon emissions at lower temperatures. This factor would cause the thermal efficiency gradient to be larger than shown in the graph, as the combustion efficiency would be decreasing, perhaps significantly, at or near the stability limit. Secondly, the design of current engines is likely unable to provide stable operation this near the ignition limited temperature. Small variations occurring on a cycle-to-cycle and cylinder-to-cylinder basis would make operation within the narrow desirable band very difficult. These cyclic variations include residual gas composition and quantity, wall, piston, head, and valve temperatures, small scale inhomogeneities, and perhaps even large-scale fluid dynamic structures that influence heat transfer rates. Therefore, while the benefits of operating near the stability limit are clear, the difficulties in attaining and maintaining operation here are significant. It seems likely that active feedback control of the intake manifold temperature, or some more easily manipulated variable, will be necessary.

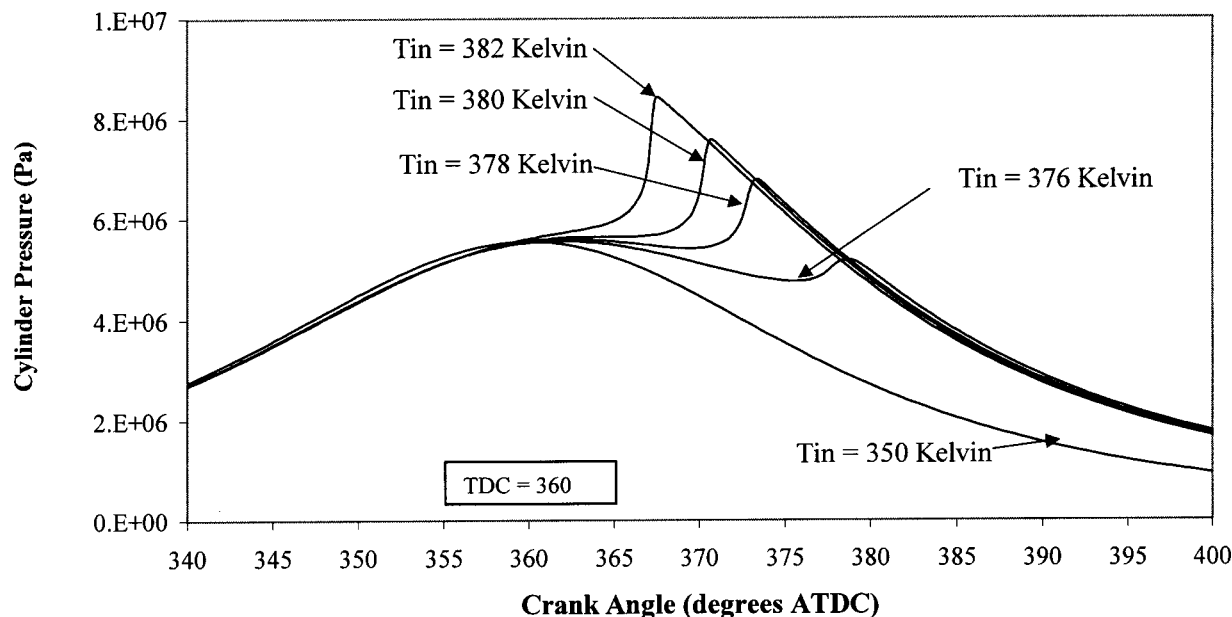


Fig. 3 Cylinder pressure over small range of inlet temperatures

Characteristic Fuels and Conditions

The above results clearly indicate the influence of natural gas composition on engine performance for two-component fuels under naturally aspirated conditions. While this is useful in terms of understanding engine performance in general, it is not representative of real world conditions where most natural gas fuel is a combination of methane, ethane, propane, and butane and where highly boosted conditions will be needed to obtain the power density required by these applications. For this reason, the simulation was run over a range of more realistic fuels and with intake boost levels of 2 bar (3 bar absolute intake pressure). The equivalence ratio and engine speed remain fixed at their previous values. These test conditions have been demonstrated experimentally, [4,5], to produce high thermal efficiency as well as high IMEP. The fuels that have been evaluated are given in Table 2, and range from pure methane (which would be typical of LNG) to a mixture with only 84% methane by volume.

Figure 4 provides information similar to Fig. 2 at the more realistic, high boost operating conditions. While the two figures are qualitatively similar, there are a number of notable differences. First, the overall indicated efficiencies are higher. This is partially due to a reduction in the importance of pumping losses (relative to engine power output) and the increased volumetric efficiency resulting from the significantly lower intake manifold temperatures. The main reason for improved thermal efficiency, though, is the lower heat losses associated with lower inlet temperature and hence lower temperatures in the cylinder throughout the cycle. These lower intake temperatures are beneficial not only in terms of efficiency but also in terms of power density and the reduced need for an outside heat source beyond the heat of compression. In general, though, the performance of the engine in terms of thermal efficiency and power ratio is very similar between natu-

rally aspirated and boosted conditions and between two component fuels and more realistic natural gas compositions.

Detailed Chemistry

The ignition and combustion of methane and natural gas fuels is relatively well understood and is discussed at length in the literature. While it is not the intention of this paper to add further detail to these arguments, some placement of kinetics within the framework of engine operation may be useful. As the simulation solves the reduced set of basic kinetic equations for natural gas combustion, the species concentrations as a function of crank angle are readily available. Figures 5 and 6 demonstrate the type of results produced by the simulation during the period of highest chemical activity.

Figure 5 provides data from which a basic understanding of the ignition event can be garnered. With only methane as a source of radical formation, the intake temperature must be quite high, note the temperature at TDC is nearly 1100 K. It can be seen from the graph that the breakdown of methane is very slow and no appreciable decrease in the methane concentration occurs until after the bulk temperature has begun to increase. The concentration of hydrogen peroxide, as described by earlier authors, [12], tends to dictate the radical formation process. The H_2O_2 concentration rises relatively late in the pre-ignition process and peaks after TDC, at which point the hydrogen peroxide begins to decompose into OH radicals which are eventually able to react with the strong C-H bond in the methane molecule. Notice also that while there are initially no higher hydrocarbons present, ethane, propane, and butane are all formed prior to ignition due primarily to the recombination of methyl radicals.

Figure 6 presents a glimpse at the physical phenomenon behind the lower ignition temperature of natural gas mixtures. The rise in hydrogen peroxide concentration begins much earlier in the process and reaches a level an order of magnitude higher than in the pure methane case, despite the nearly 50 K lower temperature. This can be attributed to the more easily decomposed higher hydrocarbons. This is particularly true in the case of butane, whose concentration has fallen by nearly an order of magnitude before a significant amount of methane has been broken down.

The criteria by which ignition after TDC can be achieved is also presented in these graphs. In both cases, even after the piston begins its decent, the heat release due to preliminary reactions is

Table 2 Natural gas recipes

Recipe	Lambda	Moles Ethane	Moles Propane	Moles Butane	Moles Methane	Mass Based Air/Fuel
A	3.333	2.50	1.00	0.50	96.00	57.198
B	3.333	5.00	2.00	1.00	92.00	56.826
C	3.333	7.50	3.00	1.50	88.00	56.489
D	3.333	10.00	4.00	2.00	84.00	56.180

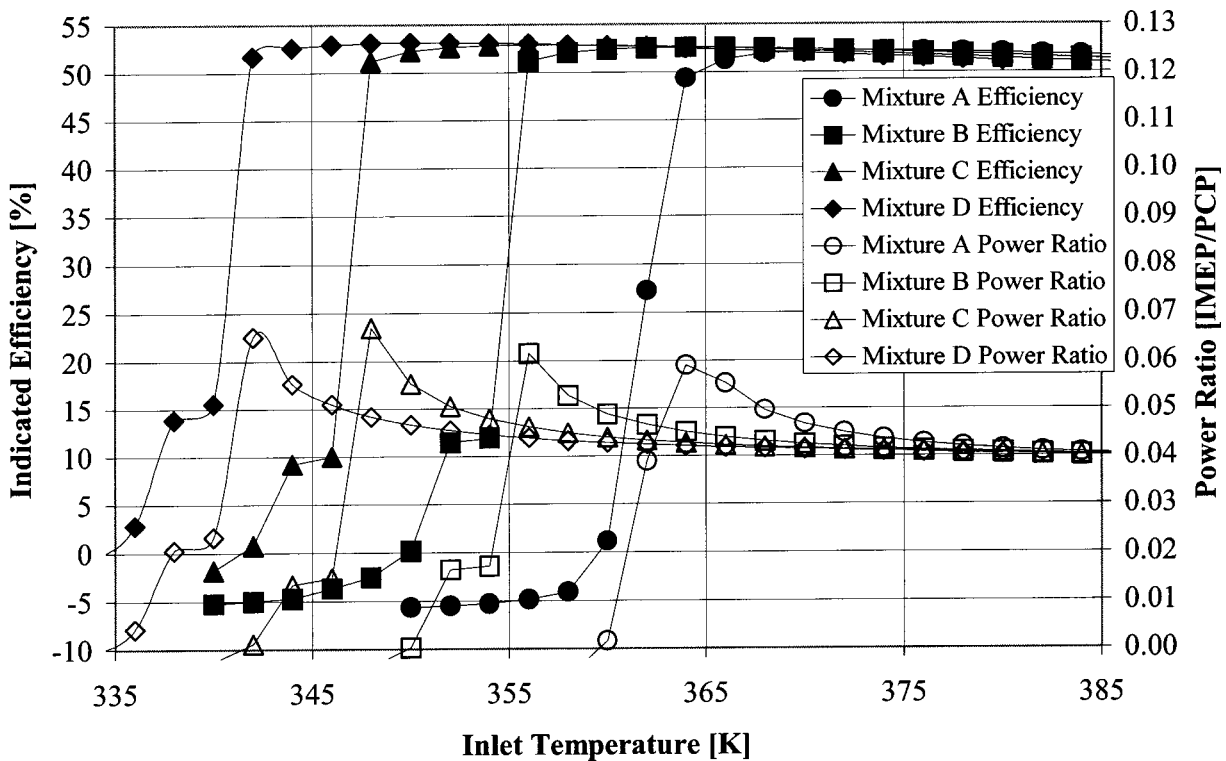


Fig. 4 Thermal efficiency and power ratio for mixtures

sufficient to result in a continued increase in cylinder temperature. That is, ignition can be made to occur after TDC, but only if heat is being released at a rate that can at least offset the temperature reducing mechanisms of volume expansion and heat transfer to the cylinder walls.

Residual Gas and Engine Stability

Much discussion in recent years has focused on the early descriptions in the HCCI literature concerning the importance of “active radicals.” While in the case of some two stroke engines with very high residual gas rates and stratification between burned gases and fresh charge some radicals could persist, it is quite clear from kinetic simulations that no “active radicals” are capable of surviving the low temperatures of the intake and early compres-

sion strokes in four stroke engines. Despite this, residual gas composition can play a role in ignition processes in some cases.

One such situation occurs in four stroke engines near the stability limit. As the intake temperature is slowly increased from an inactive state to a state of complete combustion, there is, according to the simulation, a region of unstable engine behavior. Similar cycle-to-cycle instability has also been reported experimentally in the literature, [17]. The cycle simulation presented here provides an excellent opportunity to observe this behavior. It should be noted that this behavior occurs only in a limited operating range and is not endemic to HCCI operation.

Figure 7 shows the behavior diagrammatically in some detail. Due to insufficient intake temperature during the first cycle, when no residual gases are present, complete combustion does not oc-

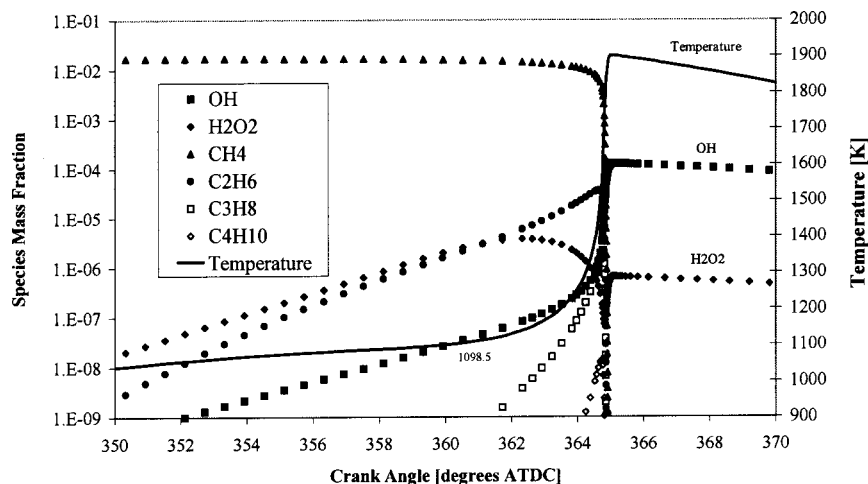


Fig. 5 Species evolution for pure methane

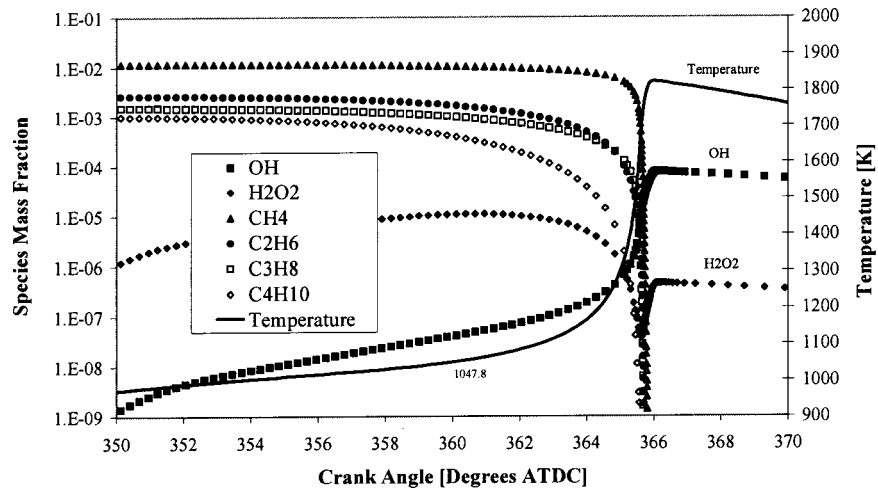


Fig. 6 Species evolution for recipe D

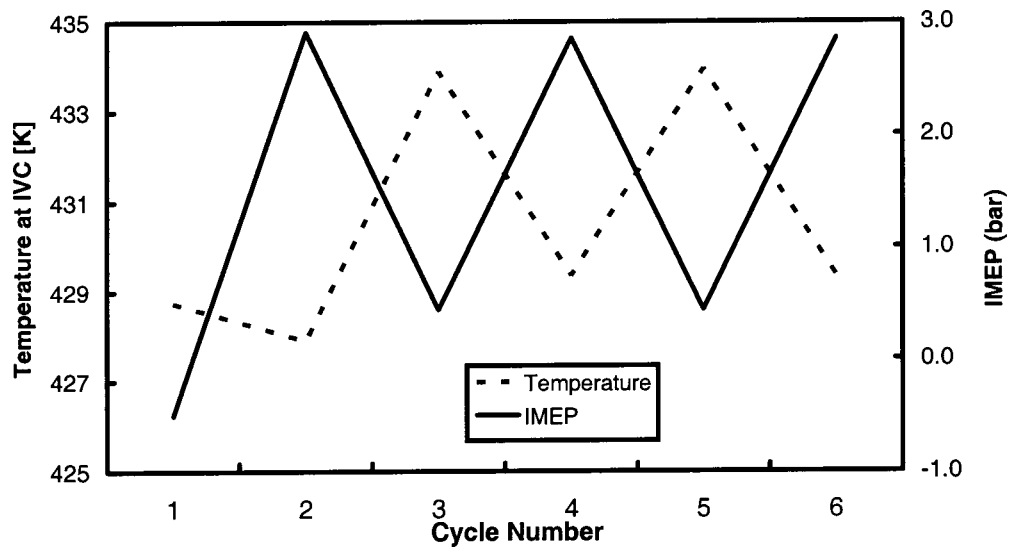


Fig. 7 Engine operation in unstable operation

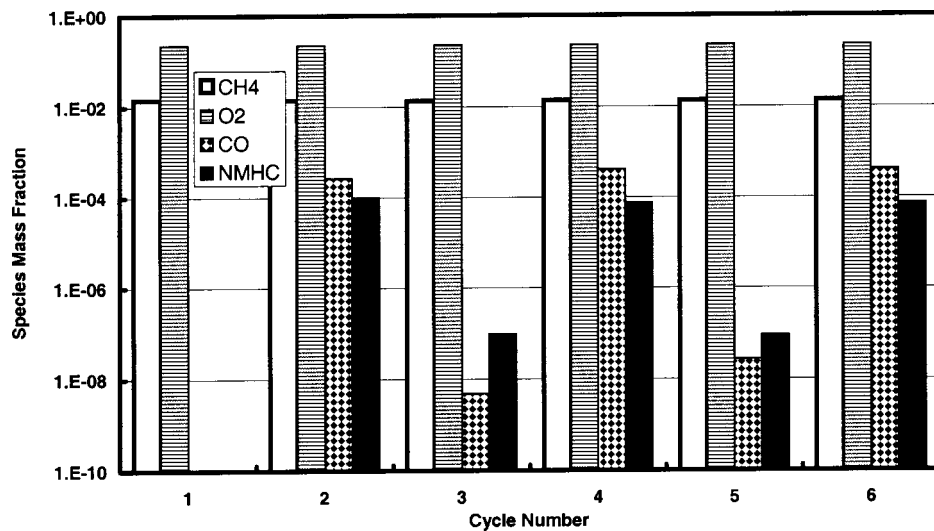


Fig. 8 Species concentration during unstable operation

cur. Despite this, there is some chemical activity and significant amounts of carbon monoxide and non-methane hydrocarbons are formed. The concentrations of these constituents on a cycle-by-cycle basis are shown in Fig. 8. During the second cycle, while the temperature of the charge is actually somewhat less than the first stroke, the presence of carbon monoxide and unburned hydrocarbons from the first cycle promote ignition and result in complete combustion of the second cycle.

Unfortunately, after this firing cycle, the composition of the residual gases is such that, despite higher temperature, the charge once again experiences only partial oxidation during the following cycle. While this mode of operation has been termed “unstable” the behavior is actually quite repeatable, with complete combustion occurring every second cycle, with only very partial oxidation on cycles between these. This cyclical operation once again demonstrates the fragile balance within the engine and the difficulty in achieving stable, repeatable operation at the very lowest allowable intake temperatures, where performance is most favorable in terms of power ratio and NO_x emissions. In the described case, in which carbon monoxide and non-methane hydrocarbons make up less than 1% of residual gas that make up less than 5% of the charge, the charge composition, rather than the initial temperature, is the primary determinate of ignition.

Conclusions

The purpose of this paper has been to demonstrate the importance of natural gas composition to the behavior of HCCI engines under conditions with high potential for commercialization. It has been shown that relatively small quantities of higher hydrocarbons (on the order of 5–15% by volume) reduce the required intake manifold temperature by up to 40 K. The simulated results also demonstrate that if stable operation can be achieved near the low temperature ignition limit, marked power density benefits can be realized. The basic chemistry leading to ignition in the cases of pure methane fuel and more realistic natural gas blends has been demonstrated for a given case. Finally, the effect of residual gas composition on combustion has shown the highly sensitive nature of operation near the ignition stability limit. The results of this study imply that market competitive HCCI engines will require precise control of any physical process that influences either mixture temperature or composition during the pre-ignition period.

Acknowledgments

The authors wish to thank the Engine Research group at Caterpillar Inc. who have provided both technical and financial support for this work. In addition, personal conversations between the lead author and researchers at U.C. Berkeley (Dr. Bob Dibble) and Lawrence Livermore National Laboratory (Dr. Ray Smith, Dr. Salvador Aceves, and Mr. Dan Flowers) contributed significantly to the understanding of ignition in HCCI natural gas engines.

Nomenclature

EGR = exhaust gas recirculation

NO_x = emissions of oxides of nitrogen (NO , NO_2 , etc.)
HCCI = homogeneous charge compression ignition
CR = compression ratio
BMEP = brake mean effective pressure
IMEP = indicated mean effective pressure
ppm = parts per million
rpm = crankshaft revolutions per minute
TDC = top dead center
power ratio = full-cycle IMEP/peak cylinder pressure

References

- [1] Ricardo, H. R., 1922, *The Internal Combustion Engine*, Vol. I, Blackie and Son Limited, London, pp. 385–414.
- [2] Kimura, S., Aoki, O., Ogawa, H., Muranaka, S., and Enomoto, Y., 1999, “New Combustion Concept for Ultra-Clean and High Efficiency Small DI Diesel Engines,” SAE Paper Number 1999-01-3681.
- [3] Christensen, M., Johansson, B., and Einewall, P., 1997, “Homogeneous Charge Compression Ignition (HCCI) Using Isooctane, Ethanol and Natural Gas—A Comparison with Spark Ignited Operation,” SAE Paper Number 972874.
- [4] Christensen, M., Johansson, B., Amneus, P., and Mauss, F., 1998, “Super-charged Homogeneous Charge Compression Ignition,” SAE Paper Number 980787.
- [5] Christensen, M., and Johansson, B., 1998, “Influence of Mixture Quality on Homogeneous Charge Compression Ignition,” SAE Paper Number 982454.
- [6] Christensen, M., Hultqvist, A., and Johansson, B., 1999, “Demonstrating the Multi Fuel Capability of a Homogeneous Charge Compression Ignition Engine With Variable Compression Ratio,” SAE Paper Number 1999-01-3679.
- [7] Christensen, M., and Johansson, B., 1999, “Homogeneous Charge Compression Ignition with Water Injection,” SAE Paper Number 1999-01-0182.
- [8] Ishibashi, Y., and Asai, M., “Improving the Exhaust Emissions of Two-Stroke Engines by Applying the Activated Radical Combustion,” SAE Paper Number 960742.
- [9] Amneus, P., Nilsson, D., Mauss, F., Christensen, M., and Johansson, B., 1998, “Homogeneous Charge Compression Ignition Engine: Experiments and Detailed Kinetic Calculations,” The Fourth International Symposium on Diagnostics and Modeling of Combustion in Internal Combustion Engines, Engine Systems Division, JSME.
- [10] Smith, J. R., Aceves, S. M., Westbrook, C., and Pitz, W., 1997, “Modeling of Homogeneous Charge Compression Ignition (HCCI) of Methane,” ASME, New York, ICE-Vol. 29-3.
- [11] Flowers, D., Aceves, S., Westbrook, C. K., Smith, J. R., and Dibble, R., 1999, “Sensitivity of Natural Gas HCCI Combustion to Fuel and Operating Parameters Using Detailed Kinetic Modeling,” ASME Fall Technical Conference.
- [12] Aceves, S. M., Smith, J. R., Westbrook, C. K., and Pitz, W. J., 1999, “Compression Ratio Effect on Methane HCCI Combustion,” ASME J. Eng. Gas Turbines Power, **121**, 569.
- [13] Aceves, S. M., Flowers, D. L., Westbrook, C. K., Smith, J. R., Pitz, W., Dibble, R., Christensen, M., and Johansson, B., 2000, “A Multi-Zone Model for Prediction of HCCI Combustion and Emissions,” SAE Paper Number 2000-01-0327.
- [14] Maigaard, P., Mauss, F., and Kraft, M., 2000, “Homogeneous Charge Compression Ignition Engine: A Simulation Study on the Effects of Inhomogeneities,” Spring Engine Technology Conference, Internal Combustion Engine Division, ASME, New York.
- [15] Fiveland, S. B., and Assanis, D. N., 2000, “A Four-Stroke Homogeneous Charge Compression Ignition Engine for Combustion and Performance Studies,” SAE Paper Number 2000-01-0332.
- [16] Flowers, D., Aceves, S., Smith, R., Torres, J., Girard, J., and Dibble, R., 2000, “HCCI in a CFR Engine: Experiments and Detailed Kinetic Modeling,” SAE Paper Number 2000-01-0328.
- [17] Ichikura, et al., 1996, “Combustion Chamber Wall Temperature and Combustion Stability,” *Engine Technology and Progress in Japan, Alternative Fuels and Engines*, Oct., Chap. 4.

N. Kyriakis

e-mail: nkyr@auth.gr

T. Manikas

P. Pistikopoulos

E. Vouitsis

Department of Mechanical Engineering,
Aristotle University of Thessaloniki,
POB 487,
541 24 Thessaloniki, Greece

Endothermic Phenomena in Diesel Particulate Filters During Regeneration

Endothermic phenomena were identified during the regeneration of the porous ceramic diesel particulate trap. These phenomena were experimentally investigated and they were found to be related to the cerium content of the fuel used for the trap loading and regeneration and to the soot mass accumulated in the trap. Carbon mass and energy balance through the regeneration process reveals that these phenomena cannot be attributed to chemical reactions in the trap. It is assumed that they are due to the effect of regeneration on the radial flow distribution through the trap, an assumption remaining to be experimentally tested. [DOI: 10.1115/1.1559905]

Introduction

Particulates emitted during engine operation can be collected in diesel particulate filters (DPFs). Depending on the filter material and construction, soot collection efficiencies as high as 95% are reported, [1–6].

The soot layers formed in DPFs, [7,8], reduce the filter overall permeability, leading thus to back-pressure increase, which in turn results in engine performance deterioration, mainly through reduction of the engine volumetric efficiency. The associated fuel consumption penalty remains well below 10% on average, reaching, however, 30% in extreme trap loadings, [9–11].

The material collected in the DPF is mainly carbon. Still, and due to the presence of sulfur in the fuel, certain quantities of sulfates are also present.

Engine operation at low load modes, due to the low exhaust gas temperature, results also in the accumulation of hydrocarbons, condensed or even adsorbed on the DPF material as well as on the soot layer, [12].

As a result of the high energy content of the exhaust gas at increased engine loads and of the presence of adequate oxygen quantity, the mass accumulated in the DPF can be auto-ignited. This oxidation process results in a partially or even completely cleaned trap (partial or complete regeneration, respectively), reducing or eliminating thus the adverse effects of increased back-pressure on engine performance.

Metal-based fuel-borne catalysts have since long been considered as a possibility to enable regeneration at relatively low exhaust gas temperatures. To this aim a number of metal-based fuel additives have been reported over the last 20 years, such as rare earths, and especially cerium, [1,2,9–11,13,14], manganese, [15,16], copper, [17], iron, [18,19], platinum, [20], and alkali metals (lithium and barium), [21], at various concentrations, ranging from a few ppm to 100 ppm, depending on the additive and the application.

In principle, regeneration is an exothermic phenomenon, since it involves burning of carbon. The heat release during regeneration, depending on the quantity of soot accumulated in the trap and on the oxidation rate, can become significant, resulting in the so-called “runaway regeneration” that may destroy (by melting or cracking) the trap. As a result, a number of trap protection techniques are reported in the literature, [1,22,23].

In cases, however, of steady-state operation of the engine,

endothermic phenomena associated with regeneration, were observed. This paper presents and investigates these phenomena and it proves that they cannot be attributed to the oxidizing activity in the trap.

Experimental Setup

The regeneration process was investigated in the experimental setup of Fig. 1.

The setup consists of

- the test engine (Daimler–Benz OM 601, 4 cylinders in line, 4 stroke, 1997 cm³, IDI).
- the engine brake (Carl Schenck W-130 eddy current dynamometer).
- thermocouples Ni–NiCr (K-type), for the measurement of temperature.
- pressure transducers (JUMO type 4 AP-30), for the measurement of pressure at trap inlet and outlet.
- exhaust gas analyzer, for the determination of the oxygen content at trap inlet (SIGNAL model 8000M).

The outputs of the measuring instruments (thermocouples, pressure transducers, exhaust analyzer, and engine speed meter) were fed to a PC through an A/D converter (National Instruments type AT–MIO–16E–10).

The trap used was the NGK type 221E (cordierite, canistered with Interam III by 3M in stainless steel housing) with dimensions 5.66”×6.00”, cell density of 100 cells/in², apparent porosity 50%, and mean pore size of 12 μm.

The exhaust pipe between engine manifold and trap was insulated, while the trap itself was exposed to the ambient air of the test cell, the temperature of which was kept roughly constant, around 30°C for all tests.

The fuel used throughout the tests was the commercially available standard diesel fuel of the Greek market, with sulfur content of 400 ppm and cetane number of 56. The fuel was blended with the EOLYS™ cerium-based fuel additive of Rhône Poulenc, at the concentrations stated accordingly.

Assuming that the additive has no effect on the quality of particulates emitted by the engine other than the presence of cerium oxidation products, and in order to eliminate the effects of soot quantity and quality on the regeneration phenomenon, the trap was loaded following a standardized loading cycle.

This loading cycle consisted of a number of different engine operation modes, including idle, enabling thus the collection of particulates of different qualities. The engine modes were selected

Contributed by the Internal Combustion Engine Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the ICE Division June 2001; final revision received by the ASME Headquarters February 2002. Associate Editor: D. Assanis.

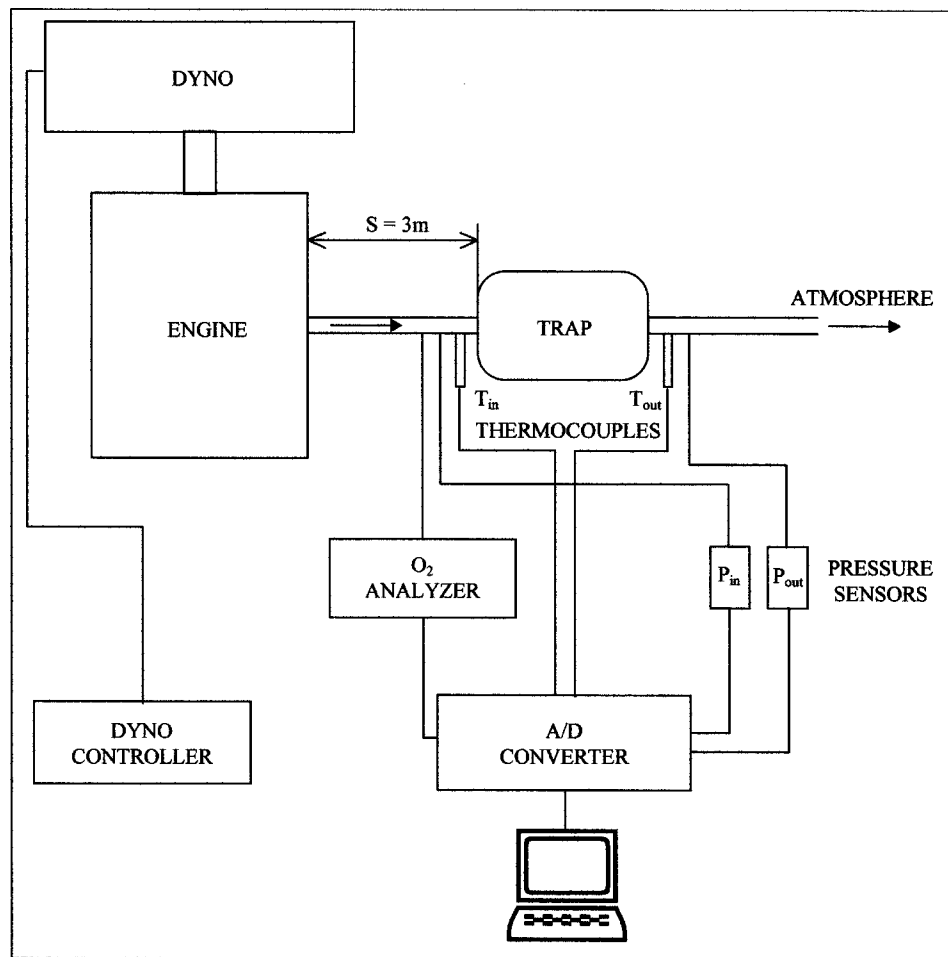


Fig. 1 Experimental setup for the study of the regeneration process

in the relatively low load range, i.e., they were associated with relatively low exhaust gas temperatures, in order to minimize the oxidation activity in the trap during loading.

The loading cycle was repeated as many times as required in order to achieve the desired trap load (integer cycles or fractions, obtained by multiplying the duration of all modes by an appropriate multiplication factor). The pressure drop across the trap at idle was used as a trap load indicator.

Results and Discussion

Figure 2 shows a typical regeneration recording of a loaded trap, with the engine operating with no additive in the fuel. As it can be seen, at the beginning of the recording the engine is idling, therefore the temperatures at trap inlet and outlet as well as the pressure drop in the trap and the engine speed present the lowest values, while the oxygen content of the exhaust is high (out of scale, at a value of about 17%).

After full throttling, the engine almost immediately reaches the speed imposed by the brake (3600 rpm) and the oxygen content of the exhaust is reduced to the stabilization value of roughly 4.2%, corresponding to the full load operation of the engine at the specific speed. The observed stabilization delay is mainly due to the length of the sampling line but also to the response time of the analyzer.

Exhaust gas temperatures at trap inlet and outlet increase slowly, due to the thermal capacity of the piping between engine and trap (it is stressed here that this piping is selected to be of significant length—see Fig. 1, i.e., of significant thermal capacity,

exactly in order to impose this slow increase of temperature. By this way the possible effect of thermocouple response is minimized).

Since the engine operates under steady-state conditions, the exhaust gas mass flow rate remains constant. Due to the increasing temperature, however, the volume flow rate through the trap increases and, as a result, the pressure drop also increases. At a certain point, however, and despite the ever increasing inlet temperature, a short period of pressure drop stabilization at a local maximum is observed. The stabilization means that the increase of

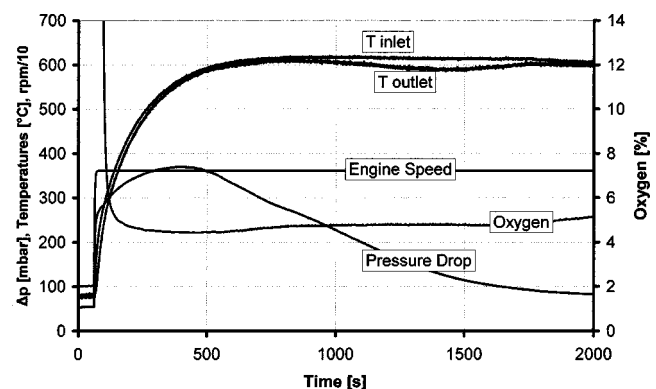


Fig. 2 Typical regeneration recording of a loaded trap—fuel without additive

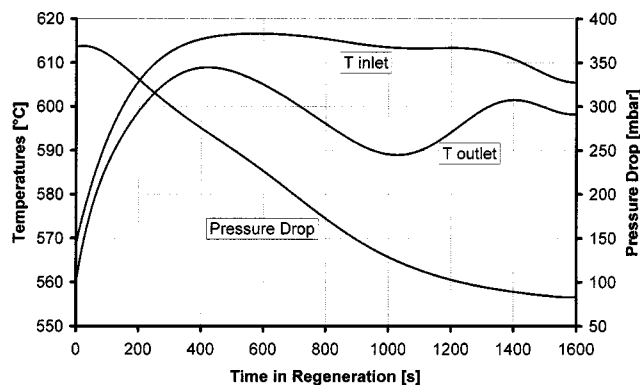


Fig. 3 Zoom in the regeneration area of the recording of Fig. 2—smoothed curves

volume flow rate is counter-balanced by a reduction of the flow resistance coefficient, i.e., by a reduction of the particulate mass in the trap. The only possible way for this reduction is the initiation of regeneration activity; therefore the inlet temperature corresponding to this maximum is defined as the regeneration temperature. As a result of the regeneration, the reduction of the flow resistance coefficient becomes more significant than the volume flow rate increase, and hence the pressure drop starts reducing, until, after about 1800 s, it reaches the stabilization value of about 100 mbar, corresponding to the clean trap under the specific volume flow rate imposed by engine speed and exhaust gas temperature. The reduction of the back pressure affects the volumetric efficiency of the engine and hence a slight increase of the oxygen content is observed—in the order of 0.2%.

Based on the above, it can be stated that the area of interest for the regeneration starts at the maximum and ends at the stabilization point of pressure drop.

Zooming in this area of the recording and after removing the electronic noise of the temperature signal, by least squares fitting, results in the picture of Fig. 3.

Figure 3 reveals that, after about 400s of the regeneration initiation, a small reduction of the reaction rate occurs, pictured by the small change of slope of the pressure drop curve. At the same time the trap outlet temperature, instead of at least following the inlet temperature or even increasing over it, as expected due to the exothermic character of the reaction, it starts decreasing for about 600s, when it starts increasing again. During this period the regeneration continues, as it is revealed by the continuous reduction of pressure drop.

The slight decrease of trap inlet temperature is attributed to the reduction of the back-pressure imposed by the trap, which results in a slight improvement of the volumetric efficiency of the engine and it is associated to the oxygen content increase already mentioned.

The reduction of trap outlet temperature could be attributed to the evaporation and/or desorption enthalpy of the condensed and/or adsorbed hydrocarbons in the trap (on the trap material itself and/or on the soot layer). Based on the relevant literature, however, the evaporation temperature is expected at significantly lower levels, below 250°C, [12], therefore this explanation cannot be accepted.

In any case, and in order to verify this statement, dedicated tests were performed, adding to the experimental setup a heated FID for the measurement of the unburned hydrocarbons downstream the trap. The results are shown in Fig. 4, and verify that hydrocarbon evaporation and/or desorption is practically completed at 250°C, well below the as above defined regeneration temperature. The peak in the “no trap” case is believed to be due to the hanging of hydrocarbons in the exhaust pipe from the engine to the sampling point. In the loaded trap cases the peaks are due to both

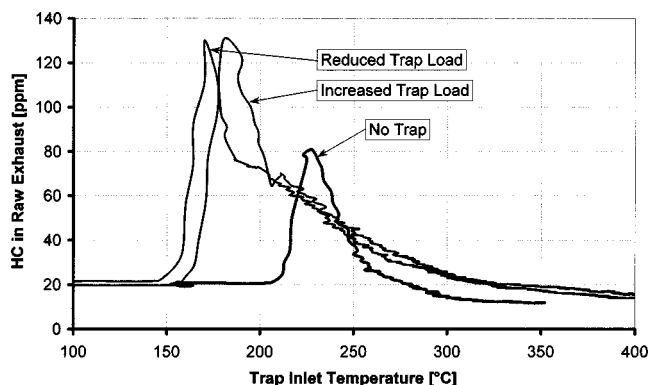


Fig. 4 Unburned hydrocarbons concentration at trap outlet as a function of trap inlet temperature

hanging and evaporation and/or desorption of hydrocarbons in the trap and in the accumulated particulates. It is stressed that in all cases the engine steadily operates at the same speed and load.

In order to verify that the observed phenomenon of temperature reduction at trap outlet is not imposed by the trap itself but it is related to the regeneration process, the “regeneration” experiment was repeated with a clean trap, i.e., with a non-preloaded trap. The resulted temperature and pressure drop time series are shown in Fig. 5.

Even though the experiment started with a completely regenerated trap, a small quantity of soot was accumulated during the transient, in terms of trap inlet temperature, operation. This small quantity of soot allows for the definition of the regeneration area, even though the recorded pressure drop values are very low, as expected.

During this regeneration, the trap outlet temperature behavior was as expected: it follows the inlet temperature, without any exothermic (the amount of soot was too small) or other excursions.

The endothermic behavior of Fig. 3 is enhanced with the presence of the cerium based catalytic additive in the fuel.

Figure 6 shows the regeneration recording when the trap was loaded and regenerated using fuel blended with 50 ppm by weight of cerium, in the form of EOLYST™ additive.

The effect of the additive on the regeneration temperature can be clearly seen: regeneration started at roughly 570°C with the virgin fuel (Fig. 3) while with 50 ppm of additive it is reduced to roughly 510°C.

The presence of the additive also increases the oxidation rate, as it can be concluded from the shorter duration of the regenera-

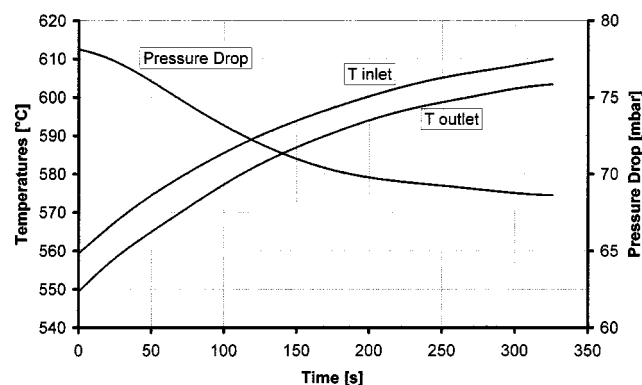


Fig. 5 Regeneration recording with clean trap—no fuel additive

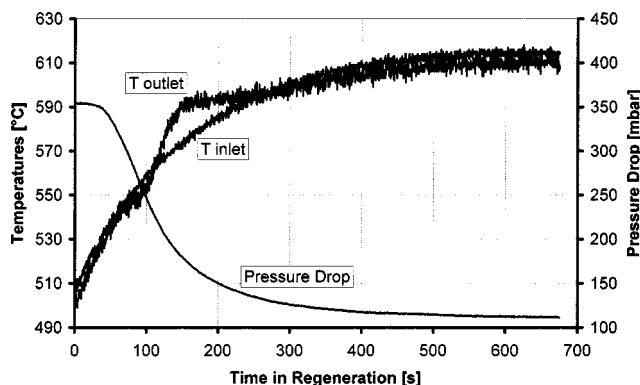


Fig. 6 Regeneration recording with trap loaded and regenerated using fuel with 50 ppm by weight of cerium

tion process, being 1600 s with virgin fuel and only 670 s in the case of 50 ppm. It is reminded that in both cases the initial trap load was practically the same.

What is of more interest, however, is the outlet temperature trace. About 70s after regeneration initiation the endothermic behavior becomes apparent, as with the virgin fuel. In this case, however, it is followed by a strong exothermic part, with the outlet temperature clearly exceeding that of the inlet, obviously due to the higher oxidation rate imposed by the additive.

Similar, but more enhanced phenomena are recorded in the case of 100 ppm by weight cerium in the fuel for trap loading and regeneration, shown in Fig. 7. The regeneration temperature is further reduced to about 495°C, the regeneration duration is reduced to 575s and the endothermic behavior is more enhanced.

Finally, even further enhancement of all the above is observed when the fuel is blended with 300 ppm of cerium, as it can be seen in Fig. 8. The regeneration temperature is determined at about 450°C, the regeneration duration does not exceed 350s and the endothermic and exothermic parts are even more obvious.

It is worth noticing that the regeneration temperature correlates very well with the inlet temperature at the point where the maximum positive difference between inlet and outlet temperature is observed.

The correlation can be seen in Fig. 9, when the same particulate mass is accumulated in the trap and different additive concentrations are used in the fuel. Figure 9 shows that the maximum effect of the endothermic phenomenon appears at lower input temperatures as the additive concentration in the fuel increases (i.e., as the regeneration temperature decreases).

Figure 10 shows the same correlation but for one additive concentration (50 ppm of cerium by weight in the fuel) and different

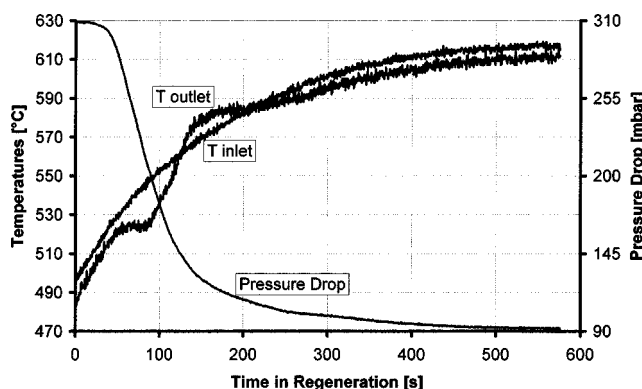


Fig. 7 Regeneration recording with trap loaded and regenerated using fuel with 100 ppm by weight of cerium

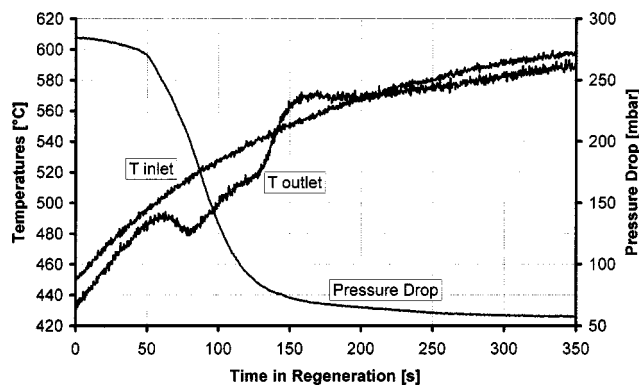


Fig. 8 Regeneration recording with trap loaded and regenerated using fuel with 300 ppm by weight of cerium

trap loads, the trap load being expressed in terms of pressure drop during idling at the end of the loading cycle. The increase of trap load results in a reduction of the inlet temperature at which the maximum effect of the endothermic phenomenon is observed.

Regeneration Calculations

In order to better understand the regeneration process, a typical regeneration recording was analyzed. The analysis consisted of

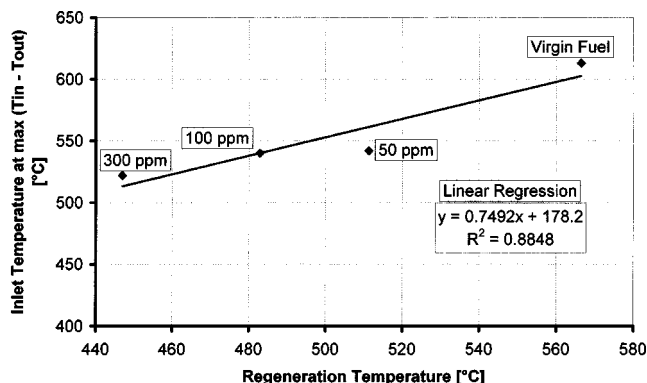


Fig. 9 The inlet temperature at maximum difference between inlet and outlet as a function of the regeneration temperature for various additive concentrations in the fuel and with the same trap load

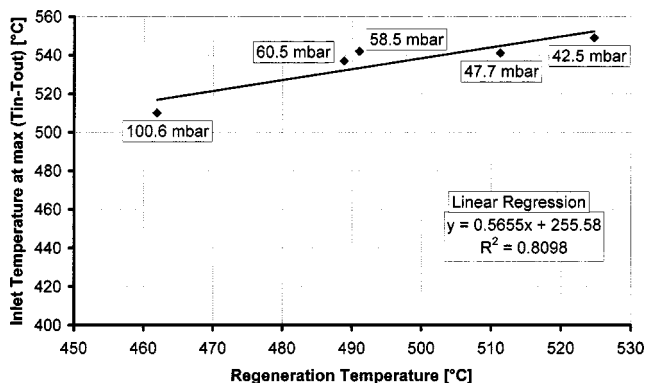


Fig. 10 The inlet temperature at maximum difference between inlet and outlet as a function of the regeneration temperature for different trap loads with a cerium concentration in the fuel of 50 ppm by weight

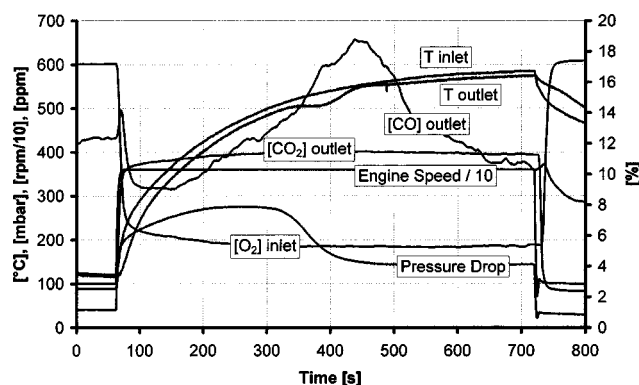


Fig. 11 Typical regeneration recording—300 ppm of cerium by weight in the fuel during both loading and regeneration

mass and energy balance calculations, based on the recorded data. The frequency of the recording was about 2 Hz (one measurement every 0.44 s).

The additive concentration in the fuel used for trap loading and regeneration in the specific case was 300 ppm by weight of cerium.

The data recorded are shown in Fig. 11. As it can be seen, the endothermic effect starts at about $t = 340$ s, and it is followed by an exothermic, starting at about $t = 390$ s. The engine speed remains constant at 3600 rpm, while carbon monoxide drastically increases during the regeneration process. Regeneration becomes visible from the stabilization of the pressure drop at the maximum value, at about $t = 360$ s, followed by the rapid reduction.

Based on the measured oxygen and on the estimated CO content of the exhaust gas at trap inlet–engine manifold (the estimation based on the stabilized area in the time interval between $t = 100$ and $t = 155$ s), the CO_2 content of the exhaust at trap inlet was calculated from the basic reaction of combustion in the cylinder of the engine. The so calculated value forms the baseline emission of the engine, therefore the integral of its difference from the measured value at trap outlet allows for the calculation of the excess carbon dioxide, produced by the combustion of the soot accumulated in the trap.

Similarly, the excess carbon monoxide produced during regeneration was calculated, and hence the total carbon oxidized in the trap during regeneration was derived.

In order to come up with the total mass accumulated in the trap, the ceria produced by the combustion of the additive has to be added, since it is practically totally collected, [2].

An independent way to calculate the trap load is through its correlation with the pressure drop (Δp) at idle under specific exhaust gas temperature.

The results of these two independent calculations are listed in Table 1, and, as it can be seen, they are in very good agreement.

In order to have the energy balance of the trap during the regeneration process, the total heat released from the production of CO_2 and CO was calculated, based on the basic chemical reactions:



and

Table 1 Measured through Δp and estimated soot mass in the trap before regeneration

Method	C in CO_2	C in CO	CeO_2	Total
	[g]	[g]	[g]	[g]
Estimated	3.85	2.16	1.80	7.81
Through Δp	-	-	-	7.40

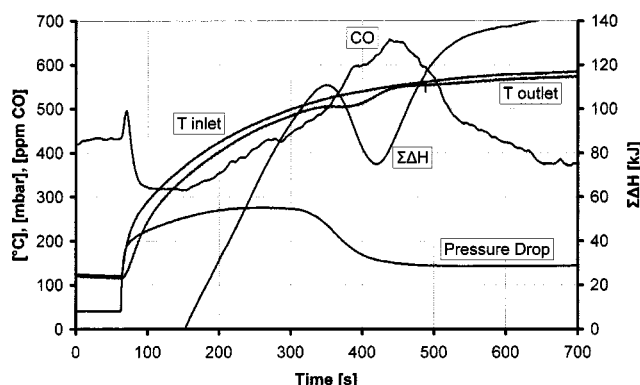
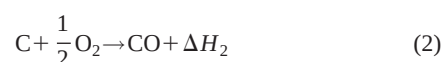


Fig. 12 The evolution of the exhaust gas enthalpy with time



where $\Delta H_1 = 32.79$ kJ/g and $\Delta H_2 = 9.21$ kJ/g, [24]. Use of the CO and CO_2 masses listed in Table 1 results in a total heat release of 146.3 kJ.

This energy was removed from the trap with the exhaust gas flow and it can be calculated from an equation of the form

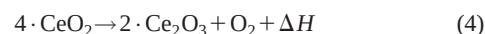
$$\sum_{\text{start}}^{\text{end}} \Delta H_i = \sum_{\text{start}}^{\text{end}} [\dot{m}_i \cdot c_{p,i} \cdot (T_{\text{out},i} - T_{\text{out,sim},i})] \quad (3)$$

where $\dot{m}_i$ is the exhaust gas mass flow rate (calculated from the engine speed), $c_{p,i}$ the thermal capacity of the exhaust gas—considered as air at the average temperature between inlet and outlet as they were measured, and $T_{\text{out},i}$ and $T_{\text{out,sim},i}$ are the measured exhaust gas temperatures at trap outlet with and without regeneration, respectively. The value of $T_{\text{out,sim},i}$ is used here as the expected trap outlet temperature without the effect of the endothermic phenomenon.

Figure 12 presents the evolution of $\Sigma \Delta H$ with time, during regeneration. Carbon monoxide trace suggests that the oxidation starts in the trap at about $t = 150$ s, but it is obviously very slow, not yet reflected on the pressure drop. The endothermic phenomenon, starting at about $t = 350$ s, is also reflected on the $\Sigma \Delta H$ curve. At about $t = 650$ s the total mass of soot in the trap has been oxidized and hence carbon monoxide and pressure drop are stabilized. The total heat removed from the trap with the exhaust gas results to 140.9 kJ, a value that is in very good agreement with the one calculated from the mass balance (146.3 kJ).

This agreement suggests that the observed endothermic phenomenon cannot be directly related to the oxidation activity in the trap. The fact, however, remains that it becomes stronger as the additive concentration increases.

An explanation was searched in the direction of temporarily heat absorption on ceria, according to the reaction



with $\Delta H = -178$ kcal/mole O_2 , [25]. In view of the ceria mass accumulated in the trap (see Table 1), however, and even if the whole mass of ceria is assumed to be transformed to Ce_2O_3 , the heat absorbed remains significantly lower than the observed. Additionally, it has to be reminded that the endotherm was visible even when there was no additive in the fuel (see Figs. 2 and 3).

It is speculated, therefore, that the observed temperature profile at trap outlet is a pseudo-endothermic phenomenon, resulting from the effect of regeneration on the radial distribution of the flow. It is anticipated in other words that the regeneration starts at the central area of the trap and decreases, due to the resulting higher temperatures, the permeability of the soot layer and of the trap material. Therefore the flow distribution shifts towards the

periphery of the trap, i.e., towards areas of lower temperature, where it is cooled, due to the increased losses to the environment. After the trap the so distributed flow is mixed again in the outlet tube, the resulting temperature of the mixture recorded as trap outlet temperature. At later stages of the regeneration, the soot layer thickness in the central areas of the trap is reduced, imposing lower flow resistance, therefore the flow returns to the original pattern (more uniform radial distribution) and the significance of the peripheral heat losses becomes less important, thus allowing the trap outlet temperature to resume the normally expected profile. The effect of cerium concentration on this outlet temperature irregularity can also be explained through the described mechanism, since higher concentrations result in more intense regenerations. The overall approach, however, remains to be tested with dedicated experiments.

Conclusions

1. Regeneration is associated with a pseudo-endothermic phenomenon, observed at the early stages of the process.
2. This pseudo-endothermic is not dependent on the trap material itself, since it is not observed with a clean trap.
3. The outlet temperature irregularity becomes stronger with the increase of cerium content of the soot and with the increase of the soot mass in the trap.
4. The inlet temperature at which the difference of the inlet and outlet temperature obtains its maximum value is well correlated to the regeneration temperature.
5. The mass balance throughout the regeneration process allowed for the estimation of the carbon mass oxidized to carbon mono- and dioxide, and hence the estimation of the total carbon accumulated and oxidized in the trap.
6. The energy balance of the regeneration process reveals that the pseudo-endothermic phenomenon is most probably a cooling effect, resulting from the radial flow distribution of the exhaust during the regeneration of the trap.

References

- [1] Pattas, K., Lee, S. H., Cho, K. R., Papadopoulos, S. G., and Kyriakis, N. A., 1994, "Ceramic Trap System for Passenger Cars," SAE Paper No. 940453.
- [2] Pattas, K., Samaras, Z., Sherwood, D., Umehara, K., Cantianni, C., Auguerre Chariol, O., Barthe, Ph., and Lemaire, J., 1992, "Cordierite Filter Durability With Cerium Fuel Additive: 100,000 km of Revenue Service in Athens," SAE Paper No. 920363.
- [3] Hoi, J., Sorenson, S., and Stobbe, P., 1995, "Thermal Loading of SiC Filters," SAE Paper No. 950151.
- [4] Bloom, R., 1995, "The Development of Fiber Wound Diesel Particulate Filter Cartridges," SAE Paper No. 950152.
- [5] Bloom, R., Brunner, R. N., and Schroeder, S. C., 1997, "Fiber Wound Diesel Particulate Filter Durability Experience With Metal Based Additives," SAE Paper No. 970180.
- [6] Ban, S., Ihara, T., Okamoto, S., Shinozaki, O., and Shinoyama, E., 1995, "Development of Porous Metal for Diesel Particulate Filter," SAE Paper No. 950739.
- [7] Pattas, K., Samaras, Z., Stamatelos, A., and Patsatzis, N., 1987, "Size Determination of the Ceramic Diesel Particulate Trap," SAE Paper No. 870253.
- [8] Pattas, K., and Samaras, Z., 1989, "Computational Simulation of the Ceramic Trap Transient Operation," SAE Paper No. 890403.
- [9] Pattas, K., Kyriakis, N., Samaras, Z., Mustel, W., and Rouveiroles, P., 1997, "The Effect of Size and Positioning of Ceramic DPFs on the Regeneration Induced by a Cerium Based Fuel Additive," SAE Paper No. 970184.
- [10] Pattas, K., Kyriakis, N., Samaras, Z., Manikas, T., Pistikopoulos, P., Mustel, W., and Rouveiroles, P., 1998, "Parametric Study of a Ceramic Diesel Particulate Trap Application on a Light Duty Truck," Int. J. Veh. Des., **20**(1-4), special issue, pp. 219-230.
- [11] Pattas, K., Kyriakis, N., Samaras, Z., Manikas, T., Mihailidis, A., Mustel, W., and Rouveiroles, P., 1998, "The Behavior of Metal DPFs at Low Temperatures in Conjunction With a Cerium Based Additive," SAE Paper No. 980543.
- [12] Lepperhoff, G., and Kroon, G., 1985, "Impact of Particulate Traps on the Hydrocarbon Fraction of Diesel Particles," SAE Paper No. 850013.
- [13] Lepperhoff, G., Luders, H., Barthe, P., and Lemaire, J., 1995, "Quasi-Continuous Particle Trap Regeneration by Cerium Additives," SAE Paper No. 950369.
- [14] Pattas, K., and Michalopoulou, C., 1992, "Catalytic Activity in the Regeneration of the Ceramic Diesel Particulate Trap," SAE Paper No. 920362.
- [15] Simon, G., and Stark, T., 1985, "Diesel Particulate Trap Regeneration Using Ceramic Wall-Flow Traps, Fuel Additives and Supplemental Electrical Igniters," SAE Paper No. 850016.
- [16] Wiedemann, B., and Neumann, K. H., 1989, "Vehicular Experience With Additives for Regeneration of Ceramic Diesel Filters," SAE Paper No. 840078.
- [17] Daly, D. T., McKinnon, D. L., Matrin, J. R., and Pavlich, D. A., 1993, "A Diesel Particulate Regeneration System Using a Copper Fuel Additive," SAE Paper No. 930131.
- [18] Terry, B., and Richards, P., 2000, "A Method for Assessing the Low Temperature Regeneration Performance of Diesel Particulate Filters and Fuel-Borne Catalysts," SAE Paper No. 2000-01-1922, 2000.
- [19] Novel-Cattin, F., Rincon, F., Trohel, O., Leflamand, M., and Dionnet, F., 2001, "Evaluation Method for Diesel Particulate Trap Regeneration Additives: Application to Five Additives," SAE Paper No. 2001-01-1914.
- [20] Jelles, S., Makkee, M., Moulijn, J., Acres, G., and Peter-Hoblyn, J., 1999, "Diesel Particulate Control: Application of an Activated Particulate Trap in Combination With Fuel Additives at an Ultra Low Dose Rate," SAE Paper No. 1999-01-0113.
- [21] Krutzsch, B., and Wenninger, G., 1992, "Effect of Sodium and Lithium Based Fuel Additives on the Regeneration Efficiency of Diesel Particulate Filters," SAE Paper No. 922188.
- [22] Pattas, K., Patsatzis, N., Michalopoulou, C., Stamatelos, A., Kikidis, P., Samaras, Z., and Sowul, H., 1988, "Ceramic Trap Regeneration Rate Control Through Bypass Technique," SAE Paper No. 880004.
- [23] Pattas, K., Samaras, Z., Kyriakis, N., Mihailidis, A., and Manikas, T., 1996, "Diesel Particulate Trap Protection Against Uncontrolled Regeneration via Selective Flow Modulation," SAE Paper No. 960128.
- [24] Smith, J. M., and Van Ness, H. C., 1987, *Introduction to Chemical Engineering Thermodynamics*, 4th Ed., Chemical Engineering Series, McGraw-Hill International Editions, New York.
- [25] Putna, E. S., Bunluesin, T., Fan, X. L., Gorte, R. J., Vohs, J. M., Lakis, R. E., and Egami, T., 1999, "Ceria Films on Zirconia Substrates: Models for Understanding Oxygen Storage Properties," Catal. Today, **50**, pp. 343-352.

Subsonic Ejector Nozzle Limiting Flow Conditions

L. J. De Chant

Research Engineer, Analytic Entrainment LLC,
1541 Summit Hills Drive, Albuquerque, NM 87112
e-mail: dechant@earthlink.net.

This paper describes an analytical method used to provide information concerning limiting flows for subsonic ejector nozzles. Three potential limiting flows have been identified and modeled using reduced control volume based analysis: (1) incipient reverse flow into the secondary inlet, (2) choked flow in the secondary inlet, and (3) choked flow in the exit mixing stream. Comparison of the methods developed here with the classical control volume portion of an ejector nozzle code have been performed and show good agreement. As such, it is concluded, that within the scope of one-dimensional control-volume based computations, that the methods developed here provide an efficient tool to help delimit the design space acceptable for ejector operation.
[DOI: 10.1115/1.1581890]

1 Introduction

Ejector-mixer nozzle systems provide important signature (acoustic and thermal) treatment technology for high-speed civil transport concepts, as well as, for civilian and military, turbofan propulsion systems. As such, the ability to delineate the underlying physically determined limits of the parameter space for subsonic ejector nozzle mixing is of considerable interest.

Generalized discussions of one-dimensional flow ejector flow modeling are provided by Addy et al. [1] and Dutton and Carroll [2,3]. Discussions with special attention to limiting conditions are provided by Emanuel [4] and Dutton and Carroll [3]. Huang and Chang [5] describe limiting flow in terms of empirical correlations for application to refrigeration systems. Similar design issues are discussed in De Chant [6,7].

The meaning of a "limiting" subsonic flow in this application describes flow fields for which the intended ejector operation; i.e., primary stream flow inducing secondary entrainment, thereby satisfying the downstream pressure constraint fails. It is possible to identify three possible limiting subsonic limiting flow situations:

1. incipient reverse flow into the secondary inlet,
2. choked flow in the secondary inlet, and
3. choked flow in the exit mixing stream.

Contributed by the Advanced Energy Systems Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the AES Division June 2000; final revision received by the ASME Headquarters December 2002. Associate Editor: G. Reistad.

These three cases are illustrated in Fig. 1.

Secondary and exit choking limiting conditions are often approached when one desires to operate an ejector in a maximum pumping condition while maintaining a subsonic exit stream. Secondary stream incipient back flow is approached during ejector start up. The analysis developed in this paper permits the computation of the upper and lower bounds on the parameter $(P_{02}/P_{01})(A_2/A_1)$. The choice of the parameter $(P_{02}/P_{01}) \times (A_2/A_1)$ was driven both by its natural appearance in the governing equations, as well as, its use in defining throttle and geometry scheduling.

2 Analysis

Analysis of the critical condition problems may be performed by considering the constant area, i.e., $A_1 + A_2 = A_3$ (a simplification introduced for clarity here, but not fundamentally required) calorically perfect, conservation equations, [1]:

$$\dot{m}_1 + \dot{m}_2 = \dot{m}_3 \quad (1)$$

$$\dot{m}_1 u_1 + p_1 A_1 + \dot{m}_2 u_2 + p_2 A_2 = \dot{m}_3 u_3 + p_3 A_3 \quad (2)$$

$$\dot{m}_1 T_{01} + \dot{m}_2 T_{02} = \dot{m}_3 T_{03} \quad (3)$$

and the exit pressure definition:

$$p_3 = p_\infty \quad (4)$$

Equations (1)–(4) can be placed into a more useful form for characterizing the critical flow conditions by rewriting them, [6]: flux ratio definition:

$$G \equiv \frac{\left[(1 + \gamma M_1^2) + \gamma M_1^2 \left(\frac{\dot{m}_2}{\dot{m}_1} \right) \left(\frac{u_2}{u_1} \right) \left\{ 1 + \frac{1}{\gamma M_2^2} \right\} \right]^2}{\gamma M_1^2 \left(1 + \frac{\gamma-1}{2} M_1^2 \right) \left(1 + \left(\frac{\dot{m}_2}{\dot{m}_1} \right) \left(\frac{T_{02}}{T_{01}} \right) \right) \left[1 + \left(\frac{\dot{m}_2}{\dot{m}_1} \right) \right]} \quad (5)$$

classical ejector mixing relationship:

$$\left[\left(1 - \frac{G}{2} \right) \gamma^2 + \frac{G\gamma}{2} \right] M_3^4 + [(2-G)\gamma] M_3^2 + 1 = 0 \quad (6)$$

and the exit pressure matching constraint:

$$\left(\frac{M_1}{M_3} \right) \left(\frac{P_{01}}{P_\infty} \right) \left(\frac{A_1}{A_3} \right) \frac{\left[1 + \left(\frac{\dot{m}_2}{\dot{m}_1} \right) \left(\frac{T_{02}}{T_{01}} \right) \right]^{1/2} \left[1 + \left(\frac{\dot{m}_2}{\dot{m}_1} \right) \right]^{1/2}}{\left(1 + \frac{\gamma-1}{2} M_3^2 \right)^{1/2} \left(1 + \frac{\gamma-1}{2} M_1^2 \right)^{(\gamma+1)/(2(\gamma-1))}} - 1 = 0. \quad (7)$$

Completing this set of equations are the definitions:

$$\frac{\dot{m}_2}{\dot{m}_1} = \left(\frac{M_2}{M_1} \right) \left(\frac{P_{02}}{P_{01}} \right) \left(\frac{A_2}{A_1} \right) \left(\frac{T_{01}}{T_{02}} \right)^{1/2} \left[\frac{1 + \frac{\gamma-1}{2} M_1^2}{1 + \frac{\gamma-1}{2} M_2^2} \right]^{(\gamma+1)/(2(\gamma-1))} \quad (8)$$

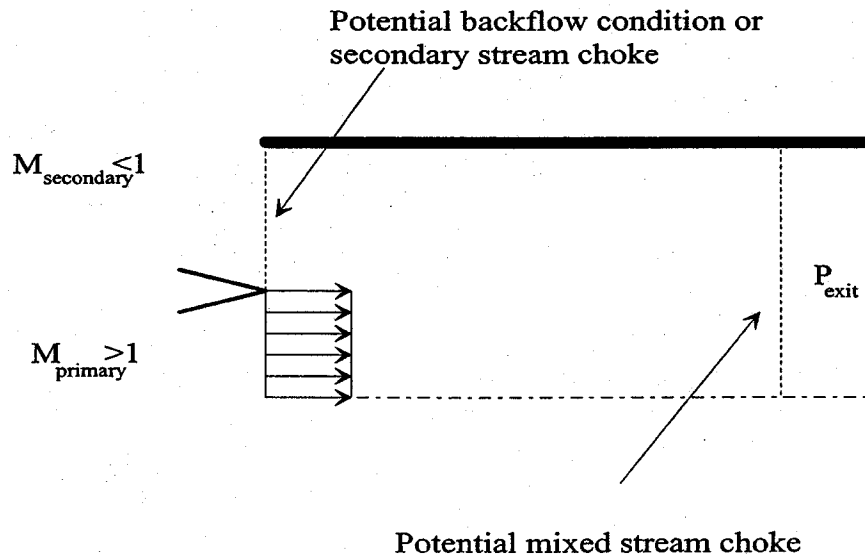


Fig. 1 A diagram showing possible limiting subsonic flow fields

$$\frac{u_2}{u_1} = \left(\frac{M_2}{M_1} \right) \left(\frac{T_{02}}{T_{01}} \right)^{1/2} \left[\frac{1 + \frac{\gamma-1}{2} M_1^2}{1 + \frac{\gamma-1}{2} M_2^2} \right]^{1/2} \quad (9)$$

Equations (5)–(9) provide the governing equations for the mixing problem in terms of the quantities of interest. As such, we may now pose the three limiting flow problems, listed in Table 1, where the parameter of greatest interest is, again, $(P_{02}/P_{01})(A_2/A_1)_{crit}$.

3 Results

In this section we compare the results of the previously described critical limit analysis to a classical, control volume based,

Table 1 Limiting flow problems

Flow Limit	Unknowns
$M_2 = 0$	$M_3, (P_{02}/P_{01})(A_2/A_1)_{crit}, G$
$M_2 = 1$	$M_3, (P_{02}/P_{01})(A_2/A_1)_{crit}, G$
$M_3 = 1$	$M_2, (P_{02}/P_{01})(A_2/A_1)_{crit}, G$

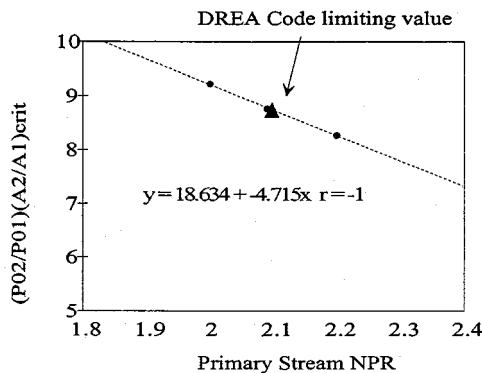


Fig. 2 Comparison of a limiting flow solution with a nonconvergent due to incipient back flow for the DREA code-based simulation. Notice that the DREA code simulation value lies on the predicted $(P_{02}/P_{01})(A_2/A_1)$ curve obtained using the current analysis.

constant area, subsonic ejector implementation, [1–3], which is part of a simulation tool: DREA; differential reduced ejector analysis, [6–9]. Note that basic control volume implementations are at the limit of their applicability, since for conditions which violate the $(P_{02}/P_{01})(A_2/A_1)_{crit}$ values they are by definition ill-posed.

3.1 Secondary Stream Incipient Backflow. Application of the control volume portion of the DREA code for a primary NPR value of approximately 2.09 ends in convergence failure. The DREA code total pressure area ratio is $(P_{02}/P_{01})(A_2/A_1) = 8.76$. It is suspected that this convergence failure is due to incipient back flow into the secondary inlet. Is this true? Solving the previous equations at various primary NPR's and comparing with the DREA code based simulation yields Fig. 2.

From Fig. 2 it is apparent that the convergence failure in the DREA code control-volume based simulation is indeed caused by incipient back flow into the secondary stream.

3.2 Secondary Stream Choking. As a test for the secondary choking constraint we consider a typical ejector problem. Conditions for this problem involve: NPR = 2.0175, $A_1/A_3 = 0.5$, $T_{02}/T_{01} = 1.0$, and $M_1 = 0.75$. A comparison of the DREA code results and Eqs. (13)–(17) results are presented in Table 2.

From Table 2 it is seen that the current method performs well at

Table 2 Comparison of current solution results to a DREA simulation for a secondary choked (saturated supersonic) flow

	DREA	Current Method	Rel. Error (DREA-Current)/Current
$(P_{02}/P_{01})(A_2/A_1)$	0.4957	0.4958	0.02%
M_2	0.9866	1.0	1.34%

Table 3 Comparison of current solution method DREA code for an exit stream choked flow

	DREA	Current Method	Rel. error (DREA-Current)/current
$(P_{02}/P_{01})(A_2/A_1)$	74.25	77.4	4.1%
M_3	0.990	1.0	1.0%

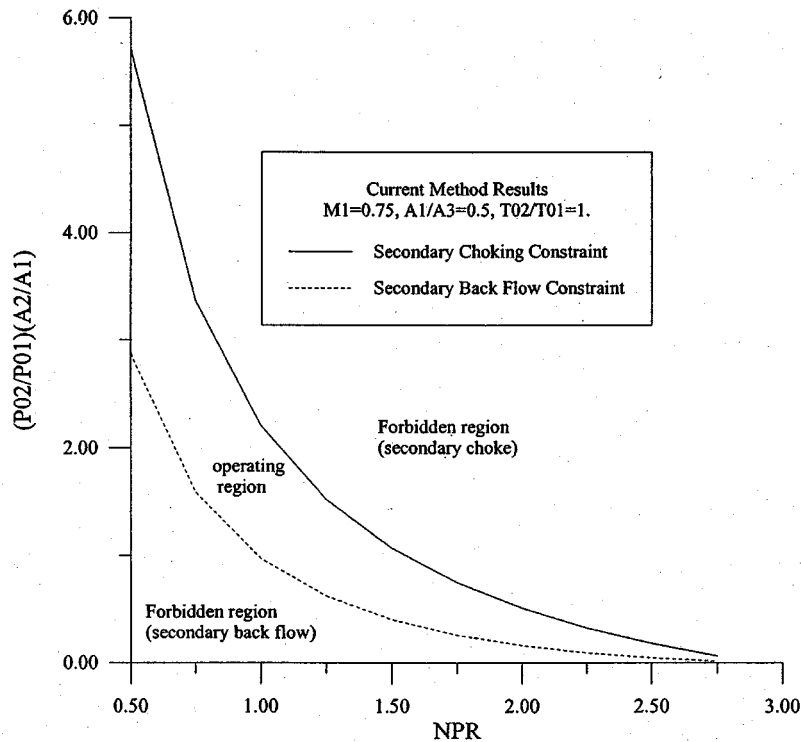


Fig. 3 Use of the current method to produce an operational map for a sub-sonic ejector

determining the conditions which cause the classical control volume implementation portion of the DREA code to choke the secondary stream.

3.3 Exit Stream Choking. Finally as a test of the exit stream choking constraint we consider another generic ejector problem. Conditions for this problem involve: $NPR = 2.5$, $A_1/A_3 = 0.01$, $T_{02}/T_{01} = 0.25$, and $M_1 = 1.0$. A comparison of DREA code results and current method results are presented in Table 3.

As shown in Table 3, the current method does a similarly good job of determining the conditions that cause the DREA code to choke the exit stream.

3.4 Application of the Current Method. In this section we consider the use of the current methods to generate a simple operational map for a generic ejector problem, $A_1/A_3 = 0.5$, $T_{02}/T_{01} = 1.0$, and $M_1 = 0.75$. Here the problem used to describe the secondary stream choke limitation was run for both secondary inlet back flow and secondary inlet choking for a range of problems and presented in Fig. 3.

4 Conclusions

This paper has sought to describe an analytical method used to provide information concerning limiting flows for subsonic ejector nozzles. Comparison with the control volume portion of the DREA code have been performed and show good agreement with the current analysis. It is our belief, that within the scope of a one-dimensional control-volume based computations that the methods developed here provide an efficient tool to help delimit the design space acceptable for ejector operation.

Acknowledgments

The first author would like to acknowledge the continued support of the Hydrodynamics Methods Group, Applied Theoretical and Computational Physics Division, Los Alamos National Laboratory, Los Alamos, NM 87545. Special thanks to J. A. Seidel and

S. B. Nadell, NASA Glenn Research Center, Cleveland OH, 44135 for help through out this project. This research was partially supported by NASA Glenn Research Center, Order C-76126-J.

Nomenclature

- A = cross-sectional area
- G = conservative flux ratio, Eq. (5)
- M = Mach number
- $\dot{m}$ = mass flow rate
- NPR = primary nozzle pressure ratio, P_{01}/p_{∞}
- P, p = pressure
- T = temperature
- u = velocity
- γ = specific heat ratio

Subscripts

- 0 = total condition
- 1,2,3 = primary, secondary, and exit stream (respectively)
- ∞ = freestream
- crit = critical (limiting) value
- exit = exit condition

References

- [1] Addy, A. L., Dutton, J. C., and Mikkelsen, C. D., 1981, "Supersonic Ejector-Diffuser Theory and Experiments," Report UILU-ENG-82-4001, Department of Mechanical and Industrial Engineering, University of Illinois at Urbana-Champaign, Urbana, IL.
- [2] Dutton, J. C., and Carroll, B. F., 1986, "Optimal Supersonic Ejector Designs," ASME J. Fluids Eng., **108**, pp. 414-420.
- [3] Dutton, J. C., and Carroll, B. F., 1988, "Limitation of Ejector Performance due to Exit Choking," ASME J. Fluids Eng., **110**, pp. 91-93.
- [4] Emanuel, G., 1982, "Comparison of One-Dimensional Solutions With Fabri Theory for Ejectors," Acta Mech., **44**, pp. 187-200.
- [5] Huang, B. J., and Chang, J. M., 1999, "Empirical Correlation for Ejector Design," Int. J. Refrig., **22**, pp. 379-388.
- [6] De Chant, L. J., 1993, "Simplified Inverse Ejector Design Tool," NASA Contractor Report, 194438.

- [7] De Chant, L. J., 1998, "Combined Numerical/Analytical Perturbation Solutions of the Navier-Stokes Equations for Aerodynamic Ejector/Mixer Nozzle Flows," NASA Contractor Report 1998-207406.
- [8] De Chant, L. J., and Andrews, M. J., 2000, "A Combined Analytical Singularity Subtraction Method and Compact Implicit Differencing Scheme for Aerodynamic Mixing," *Math. Comput. Modell.*, **31**, pp. 21–38.
- [9] De Chant, L. J., and Seidel, J. A., 2001, "Analytical Development of Non-Ideal Effects for Control Volume Based Mixer Models," *Appl. Therm. Eng.*, **21**(15), pp. 1521–1533.

Computing CHEMKIN Sensitivities Using Complex Variables

N. Butuk

e-mail: n_butuk@pvamu.edu

J.-P. Pemba

e-mail: pem@hp73.pvamu.edu

Department of Mathematics, Prairie View A&M University, Prairie View, TX 77446-4189

This paper discusses an accurate numerical approach based on complex variables for the computation of the Jacobian matrix of complex chemical reaction mechanisms. The Jacobian matrix is required in the calculation of low dimensional manifolds during kinetic chemical mechanism reduction. The approach is suitable for numerical computations of large-scale problems and is more accurate than the finite difference approach of computing Jacobians. The method is demonstrated via a nonlinear reaction mechanism for the synthesis of Bromide acid and a H_2 /Air mechanism using a modified CHEMKIN package. The Bromide mechanism consisted of five species participating in six elementary chemical reactions and the H_2 /Air mechanism consisted of 11 species and 23 reactions. In both cases it is shown that the method is superior to the finite difference approach of computing derivatives with an arbitrary computational step size h . [DOI: 10.1115/1.1469006]

1 Introduction

In developing efficient reduced chemical mechanisms for combustion simulations, it is often necessary to compute the Jacobian matrix of the system of governing ordinary differential equations. This is especially true for methods that are based on chaos theory; for example the newly introduced methods of intrinsic low dimensional manifold (ILDM) ([1]), the low dimensional center manifold (LDCM) ([2]), and other mathematical techniques ([3]). These methods all require the computation of the derivatives of the Jacobian matrix. An accurate method of determining this matrix is desired especially for large-scale problems where the analytic approach becomes tedious. The complex variables approach described in this paper and applied to a reaction mechanism for the first time is shown to be a promising method to do this.

Several approaches have been used to obtain sensitivities such as the "brute force" method, variational methods and its variants including the Green's function approach ([4,5]) and more recently the method of automatic differentiation ([6]). Recently, Shen ([7]) has developed another variation of the method that involves complex matrix manipulations. The "brute force" method is the stan-

dard approach of using finite difference. This method works well if the function is smooth. The main drawback is that the accuracy of the method is hard to analyze, and also quite unreliable for nonlinear functions. The smaller the step size the lower the truncation error (resulting from the omission of higher order terms) but the higher is the loss of significance error (resulting from subtracting two almost equal numbers, also known as "subtractive cancellation errors"). The problem is compounded further by the fact that the optimal choice of the step size is not known a priori. It may in fact depend on the independent variables themselves as well as on the type of function.

In the complex variables procedure, one simply converts the relevant code to run using complex variables. The partial derivatives in the Jacobian matrix are loaded into the imaginary part of the evaluated function and can be outputted appropriately ([8]). The application of complex variables to obtain derivatives has been described recently by Squire and Trapp ([9]). In addition, the method has been applied to obtain aerodynamic sensitivities for the Navier Stokes equations for use in aerodynamic shape optimization ([10]). It has also been demonstrated on a two-dimensional CFD Euler solver and a grid generator to obtain grid sensitivities ([11]).

This paper begins by giving a brief description of the complex variables approach in Section 2. In Section 3 we describe the procedure used to modify the CHEMKIN package to implement the complex variables approach. Section 4 will describe briefly the results of using the complex variables approach on a simple stiff bromide acid synthesis mechanism and a realistic combustion H_2 /Air mechanism using the CHEMKIN package. In Section 6 we give conclusion and recommendations.

2 Complex Variables Methods of Obtaining Derivatives Numerically

In this method, implementation is accomplished by expanding the function in a Taylor series using a complex step, then solving the resulting equation for the imaginary part and the real part. The second-order Taylor expansion of a complex function can be written as

$$f(x + ih) = f(x) + ihf'(x) + (ih)^2 f''(x)/2 + \dots \quad (1)$$

The imaginary part will contain the derivatives, and the real part will contain the function.

Table 1 Mechanism for bromide acid synthesis

$Br_2 \xrightarrow{k_1} 2Br$
$2Br \xrightarrow{k_2} Br_2$
$H_2 \xrightarrow{k_3} 2H$
$2H \xrightarrow{k_4} H_2$
$H + Br \xrightarrow{k_5} HBr$
$HBr \xrightarrow{k_6} H + Br$
$\frac{d[Br_2]}{dt} = k_2[Br]^2 - k_1[Br_2]$
$\frac{d[Br]}{dt} = 2k_1[Br_2] - 2k_2[Br]^2 + k_6[HBr] - k_5[H][Br]$
$\frac{d[H_2]}{dt} = k_4[H]^2 - k_3[H_2]$
$\frac{d[H]}{dt} = 2k_3[H_2] + 2k_4[H]^2 + k_6[HBr] - k_5[H][Br]$
$\frac{d[HBr]}{dt} = k_5[H][Br] - k_6[HBr]$
$k_1 = 9.2E-5, k_2 = 4.0E15, k_3 = 9.2E-5$
$k_4 = 4.0E15, k_5 = 1.0E15, k_6 = 1.0E-5$

Contributed by the Combustion and Fuels Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received by the CF Division March 19, 2001; final revision received by the ASME Headquarters November 1, 2001. Associate Editor: S. Gollahalli.

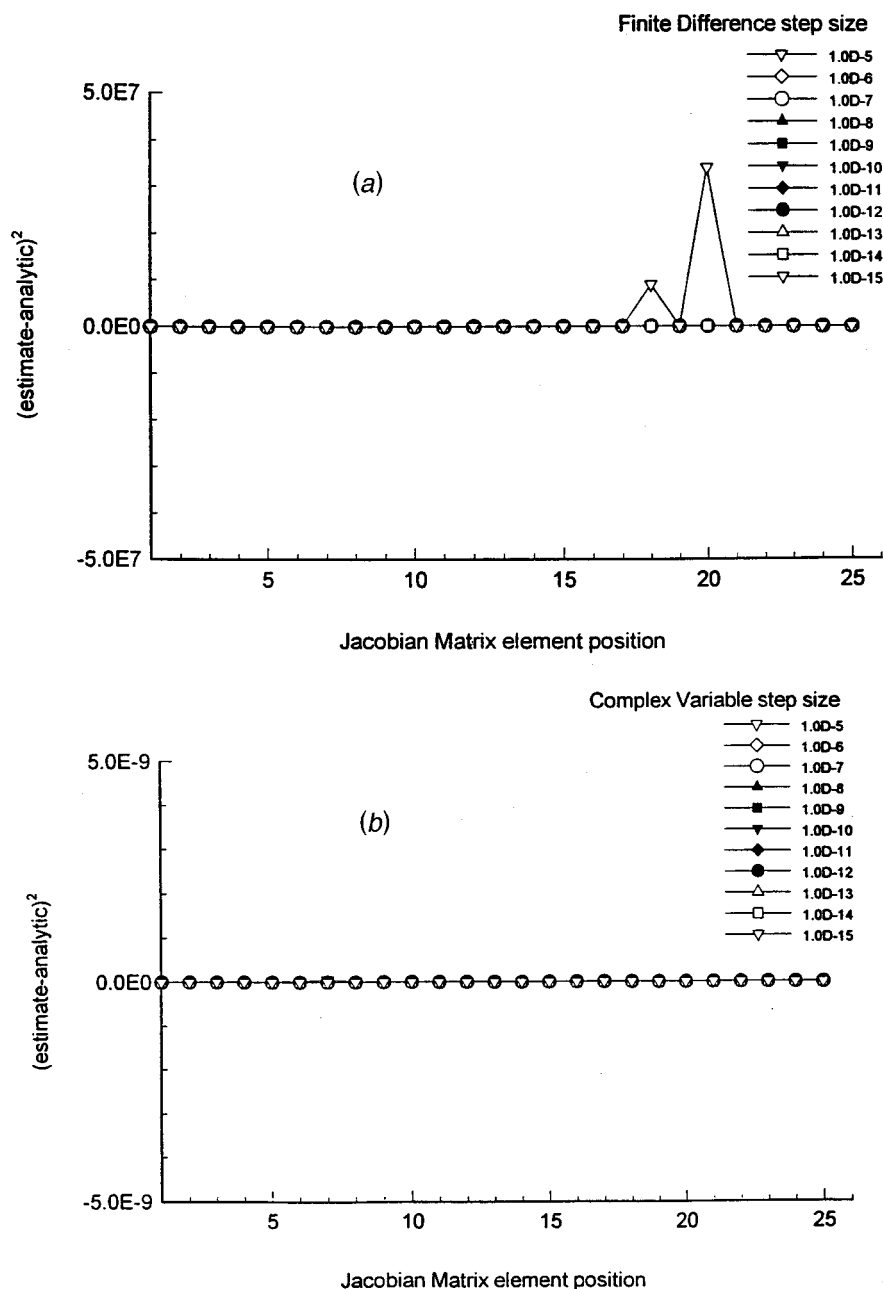


Fig. 1 (a) Finite difference computational errors of each element of the Jacobian matrix—bromine acid mechanism, (b) complex variables computational errors of each element of the Jacobian matrix—bromine acid mechanism

$$f'(x) = \text{Im}[f(x+ih)]/h \quad (2)$$

Note also that the second derivatives (and therefore the Hessian matrix) can be obtained using this approach. Pemba ([12]), has developed a procedure to do this without incurring subtractive cancellation errors. He also gives a detailed error analysis of the method. This method, therefore, simply involves evaluating the function with a complex argument such as $f = f(x+ih)$, where h is the step size and x is the independent variable. Very little modification of the software code is required. The steps to implement will be described in detail below for the CHEMKIN code.

The complex variables method is beginning to attract widespread attention and its efficiency is currently a subject of active research ([13]). Initial results indicate that when compared to the

AD method, the resulting code runs a little longer. Its main advantage, however, is the fact that it takes a shorter time to implement on a new problem.

3 Code Implementation

This section describes the modification of the CHEMKIN code to run using complex variables. We used the CHEMKIN II package for this purpose as the entire FORTRAN code was available to us. CHEMKIN is a popular software package ([14]) consisting of a collection of FORTRAN subroutine libraries for the simulation of reactive flows. The CHEMKIN package allows for the integration of the independent reactive flow equations with the specific kinetic reaction mechanism for a particular fuel.

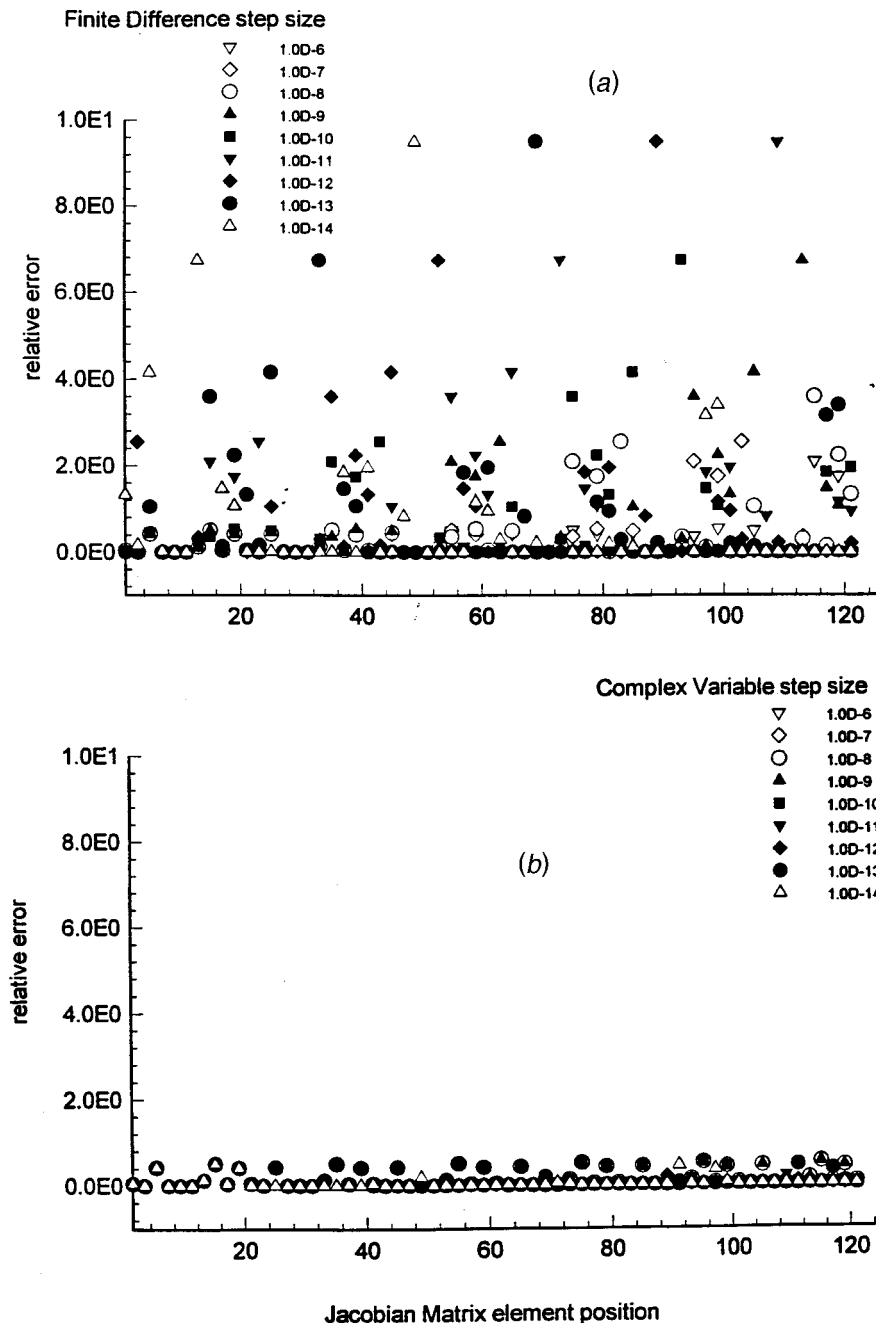


Fig. 2 (a) Finite difference computational errors of each element of the Jacobian matrix— H_2/Air mechanism (b) complex variables computational errors of each element of the Jacobian matrix— H_2/Air mechanism

In this study, the combustion of H_2/Air was simulated by a constant pressure zero-dimensional combustion model. The kinetic model consisted of 11 species and 23 reactions taken from reference ([14]). This model required the solution of the 11 species ordinary differential equations as well as the energy conservation equation for temperature evaluation, as described below. The CHEMKIN package was used to solve the equations to determine the equilibrium point. The Jacobian matrix was computed at this point by modifying a few subroutines to run using complex variables.

First of all, the functional routine used to evaluate the right-hand side of the equations at the equilibrium (fixed point) point was converted to run using complex variables. Initialization of the problem was performed as usual using the chem.bin interpreted

file output in the main calling routine. This main program evaluated the Jacobian by calling the functional routine several times. The functional routine called the necessary CHEMKIN libraries to evaluate the right-hand side of the combustion model equations. The called subroutines in the functional routine were all converted to run using complex variables. The routines included: CKRHOY, CKCPBS, CKWYP, CKHMS, and CKWT. The most challenging subroutine to convert was CKWYP, which evaluated the species production rates. It was realized that the complex mode required the real work array RCKWRK(*) to be both real and complex for some of its addresses. To avoid this complexity, we solved the problem by creating an exact identical array which was declared entirely complex and the original RCKWRK(*) array left as real

(i.e., an additional work array was created and declared complex). This solved the problem with minimal code modification. Some complex subroutines such as DMAX1 and DMIN1 functions were rewritten to output correct results. The changed subroutines were renamed and collected into one file which was then compiled and linked with the rest of the routines. The CKLIB set of routines was left intact.

4 Results and Discussions

To demonstrate the complex variables method, a simple mechanism that is stiff and nonlinear was chosen. The mechanism chosen is that for the synthesis of bromide acid shown in Table 1. This mechanism was obtained from Deco and Obradovic ([15]). The mechanism consisted of five species participating in three reversible reactions. The relevant ordinary differential equations for these species is also shown in Table 1. These equations were integrated using the VODE ODE solver of CHEMKIN. The rate constants used are shown in the table.

A point after an integration time of 40,000 seconds was chosen for the computation of the Jacobian. The finite difference (central difference) approach was used as well as the new complex variables method. Different step sizes were chosen for the computation of the Jacobians. The step sizes chosen were 1E-5, 1E-6, 1E-7, 1E-8, 1E-9, 1E-10, 1E-11, 1E-12, 1E-13, 1E-14, and 1E-15. The computations were all performed in double precision arithmetic. To compare results of the two approaches the Jacobians were also determined analytically. From the results, there was no difference in the values of the analytic matrix and complex variables matrix for all elements to within the precision used. There were some differences for some matrix elements between analytic and finite difference matrices.

The square of the errors between estimated and analytic Jacobians are plotted for each matrix element in Figs. 1(a) and 1(b). Figures 1(a) show the finite difference errors and Fig. 1(b) the complex variable errors. In Fig. 1(a), using finite difference, it is clear that the errors can be quite significant for some elements of the Jacobians matrix. The magnitudes of the errors can be as high as 4E7 as the step size is reduced. On the other hand in Fig. 1(b), the errors for the complex variables approach are negligible (never more than 1E-9) for all step sizes and for all matrix elements. This clearly shows that the accuracy of the method for chemical kinetics can be quite good. To within machine precision, therefore, the complex variables approach of computing Jacobians is exact, to order $O(h^2)$ where h is the step size. Next is discussed the results obtained using a realistic combustion mechanism with CHEMKIN simulation software.

The hydrogen/air reaction mechanism chosen consisted of 11 species and 23 elementary reversible reactions ([14]). The species were: $\mathbf{Y} = [\text{H}_2, \text{H}, \text{O}_2, \text{O}, \text{OH}, \text{HO}_2, \text{H}_2\text{O}_2, \text{H}_2\text{O}, \text{N}, \text{N}_2, \text{NO}]^T$. The combustion of H_2/Air was simulated by a constant pressure zero-dimensional combustion model. This model required the solution of the species ordinary differential equation as well as the energy conservation equation for temperature evaluation. These equations are

$$dY_k/dt = \omega_k W_k / \rho, \quad k = 1 \dots n \quad (3)$$

for n species and the energy equation

$$dT/dt = 1/\rho c_p \sum h_i \omega_i W_i, \quad i = 1 \dots n \quad (4)$$

where Y_k is the mass fraction of k th species, ω is the molar production rate of the species, W is the molecular weight of the species, ρ is the density of mixture, T is the temperature, c_p is the mean specific heat of mixture at constant pressure, and h_i is the specific enthalpy of species i . The energy equation was converted to integral form and solved iteratively for temperature assuming constant enthalpy (this, however, is not necessary and was done only to reduce the number of ordinary differential equations to

solve). This allowed for the integration of species ordinary differential equations only in which the species concentrations were of the same order of magnitude.

The Jacobian matrix was computed at the equilibrium point, using both finite differencing and the new complex variables method (implemented using the new modified CHEMKIN routines as described above). The central differencing formula was used to compute the finite difference Jacobians. The Jacobians were computed using the same step sizes as for bromide acid except for 1E-5 and 1E-15 which were not used.

Since the analytic Jacobian was not available for this case, we picked one computed Jacobian using an h value of 1E-8 as a reference Jacobian. At this h value, the complex variable Jacobians were very close to the finite difference Jacobians. We then used this reference Jacobian to compute the deviations of finite difference and complex variable Jacobian computations using the different step sizes (note that an accurate method should not be affected by the step size used). Relative errors (a more accurate term could be deviations) were computed based on this reference Jacobian and are shown plotted in Figs. 2(a) and 2(b). Figure 2(a) shows the relative deviations for the finite difference computations. From this plot, it is clear that the variations (or spread) of the deviations was quite large and the data highly scattered. On the other hand, the computations performed using the complex variable approach and shown in Fig. 2(b) indicate a very small spread. These results indicate clearly the influence of step size h , in computing the Jacobians. Finite difference computations are unstable and can vary widely; whereas the complex variable computations have a very small variation and are practically step size independent. The effect of varying the step size is dramatic using the finite difference approach but is within the machine precision using the complex variable approach. To shed further light on the results, we computed the standard deviations of both data sets. This was done on the data generated by all values of step sizes used. The standard deviation of the finite difference approach was 28.14 and the standard deviation of the complex variable approach was 0.889. This showed a clear distinction between the two approaches and the step size independence of the CV approach.

5 Conclusions and Recommendation

A description of the new complex variables approach of computing sensitivities has been given. The method was implemented and compared with the traditional finite difference approach on two sample mechanisms; a simple bromide acid synthesis mechanism and a realistic hydrogen/air combustion mechanism. The modification of the CHEMKIN software to implement the complex variables approach has been outlined. It has been shown that this new complex variables approach is step size independent and is more accurate than the traditional finite difference approach. To implement this approach requires minimum modifications of the necessary computational routines. In order to apply the system to large-scale problems, modification of the necessary computational software such as CHEMKIN routines has been demonstrated.

Acknowledgments

This work has been partly supported by Prairie View A&M University Research Enhancement Program and by DOE under grant number DE-FG26-00NT-40830.

References

- [1] Maas, U., and Pope, S. B., 1992, "Simplifying Chemical Kinetics: Intrinsic Low-Dimensional Manifolds in Composition Space," *Combust. Flame*, **88**, pp. 239–264.
- [2] Butuk, N. K., and Pembra, J.-P., 2001, "Computing Low Dimensional Invariant Manifolds of Reaction Mechanism Using Genetic Algorithms," *AIAA Paper* 2001-1076.
- [3] Tomlin, A. S., Turanyi, T., and Pilling, M. J., 1997, "Mathematical Tools for the Construction, Investigation and Reduction of Combustion Mechanisms," *Low Temperature Combustion and Autoignition*, M. J. Pilling and G. Hancock, eds., Elsevier, New York, pp. 293–437.

- [4] Rabitz, H., Kramer, M., and Dacol, D., 1983, "Sensitivity Analysis in Chemical Kinetics," *Annu. Rev. Phys. Chem.*, **34**, pp. 419–461.
- [5] Hwang, J.-T., Dougherty, E. P., Rabitz, S., and Rabitz, H., 1987, "The Green's Function Method of Sensitivity Analysis in Chemical Kinetics," *J. Chem. Phys.*, **69**(11), pp. 5180–5191.
- [6] Carmichael, G. R., Sandu, A., and Potra, F. A., 1997, "Sensitivity Analysis for Atmospheric Chemistry Models via Automatic Differentiation," *Atmos. Environ.*, **31**(3), pp. 475–489.
- [7] Shen, J., 1999, "A Direct Method of Calculation Sensitivity Coefficients of Chemical Kinetics," *J. Chem. Phys.*, **111**(6), pp. 7209–7214.
- [8] Newman, III, J. C., Anderson, K. W., and Whitfield, D. L., 1998, "Multi-disciplinary Sensitivity Derivatives Using Complex Variables," Computational Fluid Dynamics Laboratory, Mississippi State University, MS, MSSU-COE-ERC-98-08.
- [9] Squire, W., and Trapp, G., 1998, "Using Complex Variables to Estimate Derivatives of Real Functions," *SIAM Rev.*, **10**(1), pp. 110–112.
- [10] Anderson, K., Nielson, E., Whitefield, D., and Newman, J., 1999, "Sensitivity Analysis for the Navier-Stokes Equations on Unstructured Meshes Using Complex Variables," AIAA Paper 99-3294.
- [11] Ye, F., Butuk, N. K., Huque, Z., and Biney, P., 1999, "CFD Optimization of an Integrated Multi-Tube Pulse Detonation Engine Inlet Concept Using Complex Variables," AIAA Paper 99-3229.
- [12] Pemba, J.-P., 2001, "The Complex Method in Numerical Differentiation," preprint, submitted for publication.
- [13] Martins, J. R. R. A., Sturdza, P., and Alonso, J. J., 2001, "The Connection Between the Complex-Step Derivative Approximation and Algorithmic Differentiation," AIAA Paper 2001-0921.
- [14] Kee, R. J., Rupley, F. M., Meeks, E., and Miller, J. A., 1996, *Chemkin III User Manual*, Reaction Design, San Diego, CA.
- [15] Deco, G., and Obradovic, D., 1996, *An Information-Theoretic Approach to Neural Computing*, Springer-Verlag, New York, pp. 152–156.