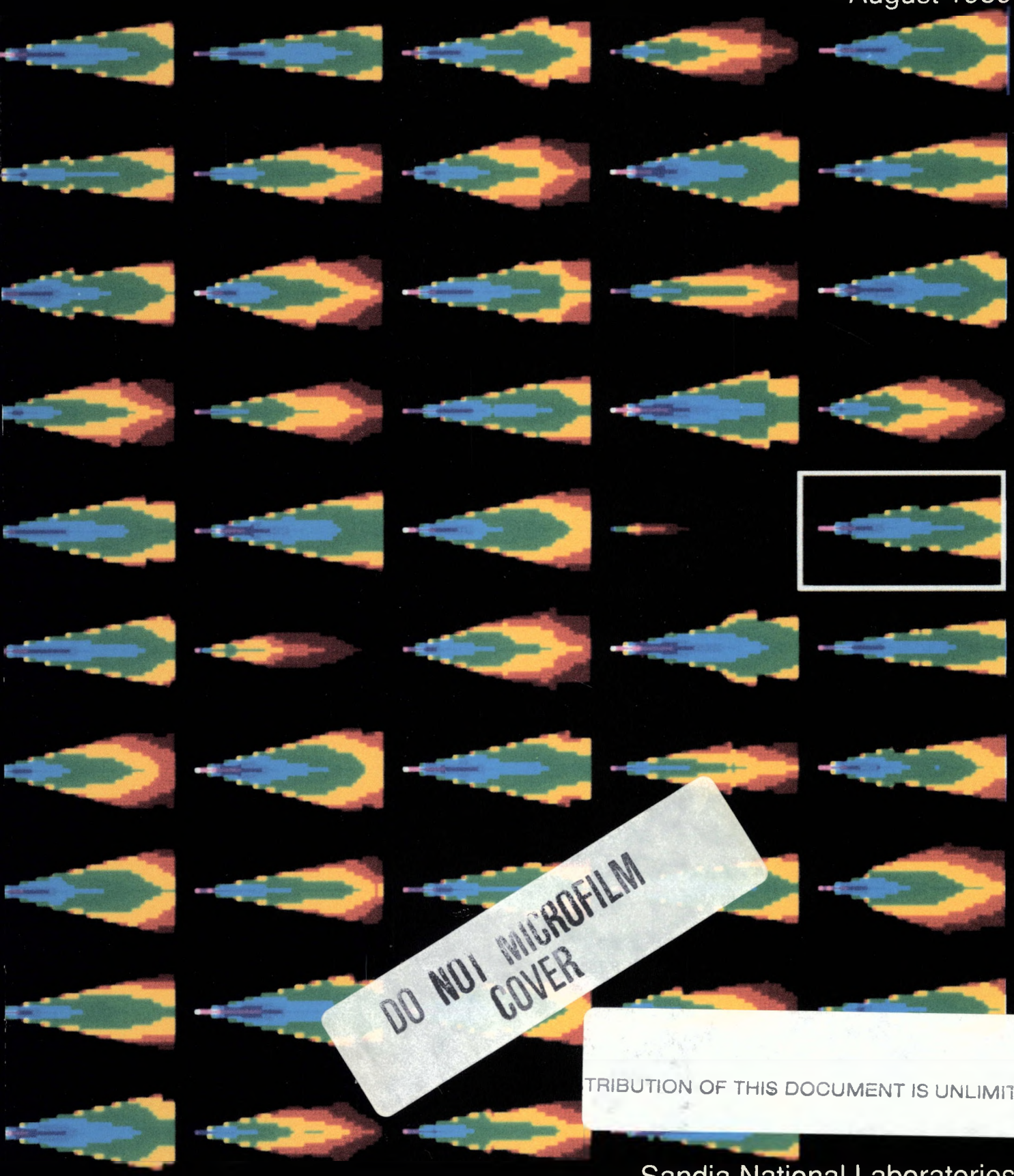
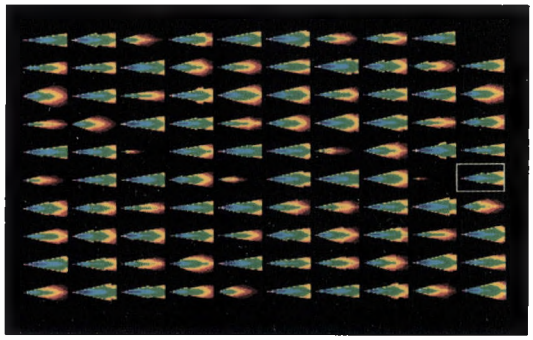
SANDIA TECHNOLOGY

August 1989



Sandia National Laboratories
Albuquerque, New Mexico • Livermore, California

Cover Photo:
A Composite High-Altitude Radiation Model computer program generated these 99 images for a simulated altitude of 150 km. Six parameters were varied systematically to produce the 98 images surrounding the nominal-value image, enclosed in the white box. The article beginning on Page 2 describes a sensitivity analysis performed on the significant model parameters.



Management Staff
Arlyn Blackwell

Special Publications Staff
Managing Editor
Randall C. Maydew

Supervising Editor
Ann L. Hogan

Lead Scientific Editor
Hal Pruett

Scientific Editors
Joseph F. Calek
Bruce C. Dale
Jim Leonard

Graphic Designer
Nancy M. Hunt

Art Director
Janet L. Jenkins

Artist
John L. Daniel

Publication Coordinator
Debbie Johnson
(505)846-3839

Published by
Sandia National Laboratories

Sandia is a multiprogram laboratory operated by AT&T for the Department of Energy with major facilities at Albuquerque, New Mexico, and Livermore, California, and a test range near Tonopah, Nevada. We have major research and development responsibilities for nuclear weapons, arms control, energy, environment, and other areas of strategic importance to national security. Our principal mission is to support national defense policies by ensuring that the nuclear weapon stockpile meets the highest standards of safety, security, control, and military performance. Selected technical activities and accomplishments are reported in two corporate publications. Unclassified articles appear in ***Sandia Technology***. Classified work is reported in ***Sandia Weapon Review***.

SAND--88-9025

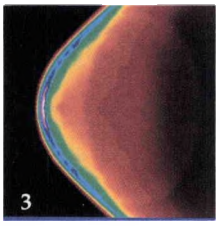
DE90 005692

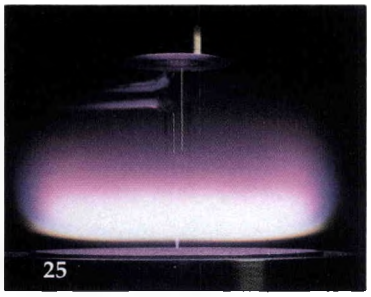
SAND 88–9025
Vol. 13, No. 3

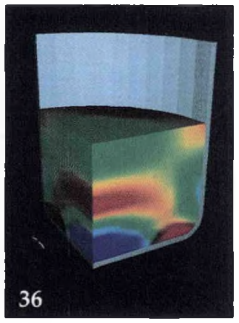
SANDIA TECHNOLOGY

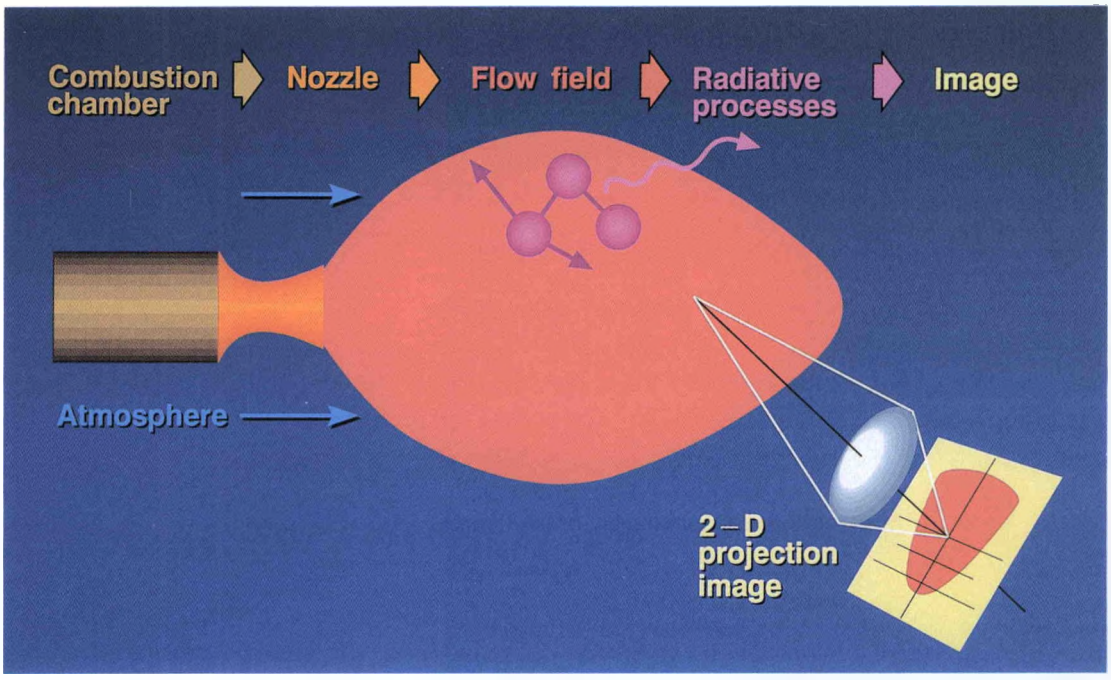
August 1989

CONTENTS









Sensitivity Analysis of Computed Rocket Plume Signatures

The characterization of variability in rocket plume signatures is accomplished using formal sensitivity analysis techniques applied to large computer codes.

Exhaust gases from rocket engines are bright sources of infrared energy. Sensors in surveillance and strategic defense systems view these infrared signatures. At Sandia, we are providing input into system designs by examining the expected variability of rocket plume signatures; we apply sensitivity analysis techniques to the computer models used to predict such signatures.

Computer models of rocket plumes along a trajectory are used to develop the control algorithms for elements of the Strategic Defense System. These system elements will use the spectral signatures arising from the rocket exhaust for surveillance, target acquisition, tracking, and kill assessment. Accurate plume signature models are necessary to develop reliable localization algorithms and hardware for strategic defense systems. Currently, the extent and quality of available plume signature data is insufficient to provide comprehensive verification of the plume signature codes. Differences among various missile systems, altitude regimes, wavelengths, and engagements of interest are not yet fully understood. Moreover, significant uncertainties exist in many of the physical parameters that describe potential threat boosters and in the physics embodied in the plume models themselves.

In work supported by the Air Force Astronautics Laboratory (Edwards Air Force Base, CA), we are systematically examining the sensitivities of the standard plume signature codes. Over the years industrial contractors have developed and refined plume signature codes under the auspices of the Joint Army, Navy, NASA, Air Force (JANNAF) committees and more recently, the Strategic Defense Initiative Office (SDIO). Periodically, "standard" code versions are released that designers use to predict plume phenomena in sensor and system design programs.

Figure 1.
Numerical computer models of rocket plumes include models of a number of phenomena. These include the fluid mechanics and chemistry occurring in the combustion chamber, nozzle, and the environment into which the gases are exhausted.

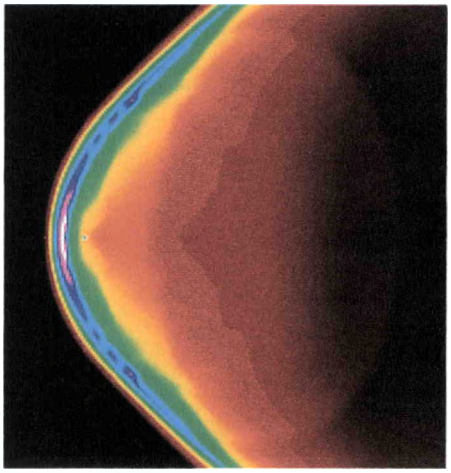


Figure 2.
This photograph illustrates the computed infrared plume signature of a large liquid-fueled rocket near 150 km altitude viewed from an aspect angle of 45° off the nose of the rocket. The spatial dimensions illustrated are roughly 1 km.

During the boost phase, rocket plumes are observable infrared sources for surveillance and targeting.

Computer models of high altitude rocket plumes mathematically describe a number of processes as illustrated in Figure 1. Fuel in the combustion chamber burns, and the energetic exhaust products are expanded through a nozzle and exhausted out of the rocket. Radiative emission from hot, vibrationally excited molecules gives rise to the infrared signature. Figure 2 illustrates the predicted signature for a large liquid-fueled rocket at 150 km altitude. This plume is shown for a viewing aspect angle of 45° off the nose of the rocket. The small blue patch in the yellow region of the plume is the vacuum core, located immediately behind the rocket; the plume is quite large and surrounds the missile. The spatial dimensions illustrated are roughly 1 km. The missile itself is roughly 1 m in diameter and 10 m in length.

The plume code used to generate this plume image is complex and includes hundreds of physical parameters whose precise values are not well known. Understanding the effect of these parameter uncertainties on plume model predictions is a necessary ingredient of careful model use by systems designers and, further, is required for continuing model development and improvement.

As Figure 3 illustrates, the observed plumes are dependent on the altitude of the rocket. The plume is the most compact under exoatmospheric conditions, and consists of the exhaust gases themselves as they expand into a vacuum. The characteristic size of a plume in this region is comparable to the rocket booster itself, on the order of 1 to 10 m.

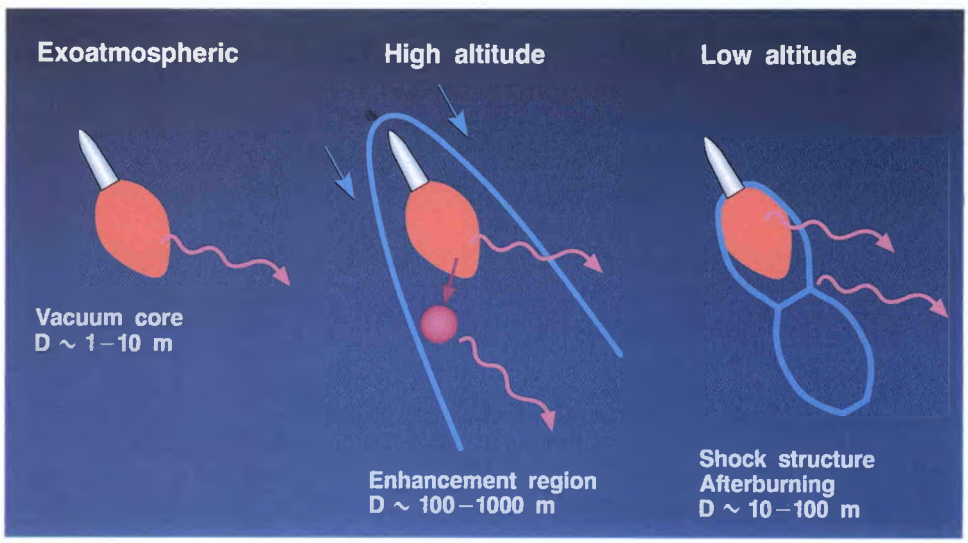


Figure 3.
Interactions between the rocket exhaust gases and the atmosphere depend on the altitude. The characteristic dimension (D) of the rocket plume also varies depending on the altitude regime.

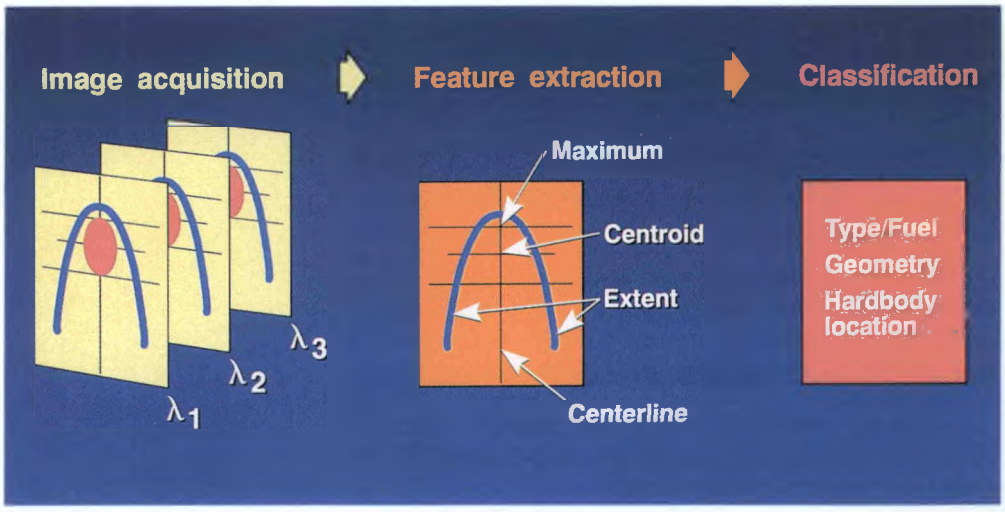


Figure 4. Optical sensors capture a two-dimensional image of the rocket plume in a particular spectral region. Image features extracted from the plume image are used for classification.

The region between 70 and 300 km is known as the high altitude regime, and the plume signature arises from two distinct effects—the exhaust gases near the rocket itself, referred to as the intrinsic or vacuum core and a larger, spatially distinct region known as the enhancement region. The enhancement region results from the interaction of plume gases with the atmosphere. These plumes are quite large, from 100 to 1000 meters, and as Figure 3 shows, they often surround the rocket. This regime is a challenging one for missions that seek to identify the missile hardbody within the spatially extended plume (see Figure 2).

At low altitudes, up to roughly 70 km, the plume consists of the exhaust gases and a complex shock structure which arises from interactions with the relatively dense atmosphere. Also contributing to the plume signature at low altitudes is afterburning— combustion of unburned fuel fragments in the presence of atmospheric oxygen. These effects lead to plume sizes which are roughly 10 to 100 m.

Image features are extracted from the plume images for use in system applications. Signatures in a number of spectral wavelength regions may be used. Figure 4 illustrates typical image features; they include the total intensity, the location of the image maximum, the location of the radiance centroid, and the spatial extent. These extracted image features may be used in rocket typing and hardbody localization. The plume images obtained by system sensors are projections of the plume in the direction of view. The images have a very strong aspect angle dependence; knowledge of the aspect angle is required to deconvolve the images.

We have focused our attention on the sensitivity of a high altitude rocket plume model developed under the sponsorship of the Air Force Astronautics Laboratory. This model is called the Composite High Altitude Radiation Model, or CHARM.

Parameter uncertainties affect the output precision of complex physical models.

Figure 5 illustrates graphically the sensitivity analysis for one output feature f and two parameters, p_1 and p_2, used in the model calculation of f. The nominal value of the output feature, computed from the model, is designated f^0. The objective is to understand the feature solution surface, as illustrated, for various values of the model parameters p_1 and p_2. Because the parameters are not precisely known, sensitivity analysis attempts to understand the shape of the entire response surface for plausible values of the parameters.

Two factors are important in a sensitivity assessment: the level of uncertainty in a parameter as well as the sensitivity of the model prediction to that parameter. A parameter may have a large uncertainty associated with its value but have little influence on the solution, while even a small uncertainty in a parameter to which the solution is very sensitive could be highly influential. Furthermore, the models approximate complex physical phenomena and differ in their approximations, so that the model of choice itself is "uncertain."

An uncertainty distribution is illustrated for parameter p_1. Such distributions must be defined for

all parameters of interest. The interaction of parameters when both (or several) are away from their nominal values cannot be evaluated by simply varying a single parameter and evaluating the slope; more precise searching techniques are required.

A variety of techniques may be employed to investigate the relationship parametric variations in input and model parameters and the output features. Our sensitivity analysis approach uses a number of elements including the application of formal sensitivity techniques. Two global techniques that examine the response surface (in contrast to local techniques that determine the slope at the nominal point) have been applied to this problem. In general, these techniques sample the parameter space to examine influences on feature values (see Box A).

We have applied two formal sensitivity techniques to rocket plume calculations.

We have applied the two viable techniques illustrated in Box A: the Fourier Amplitude Sensitivity Test (FAST) and Latin Hypercube Sampling (LHS). The first technique–FAST–was developed and has been applied in the chemical kinetics community. In FAST, each uncertain parameter is arbitrarily associated with a specific Fourier frequency. The system sensitivities are then determined by executing the model for discrete values of the Fourier transform variable and computing the Fourier coefficients of the model outputs. The basic sensitivity measure in FAST is the partial variance; it is the fraction of the total variance of a model output due to the variation of a particular parameter. The partial variance is computed by summing the Fourier coefficients at the harmonics of the parameter frequency. Partial variances associated with

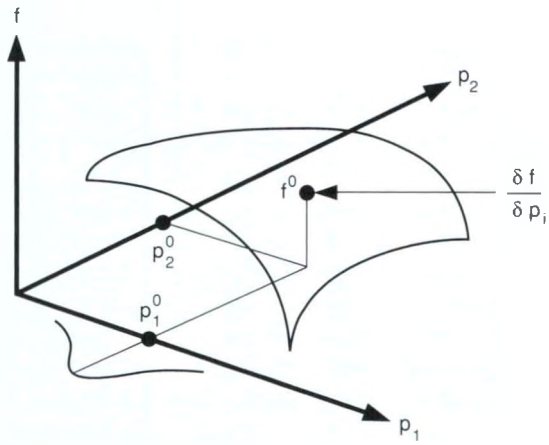


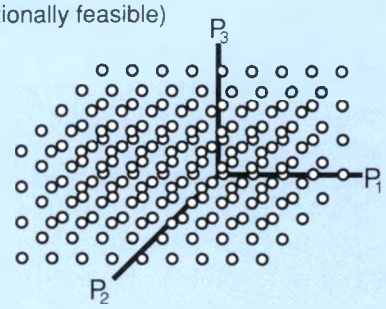
Figure 5. Sensitivity analysis examines the solution surface of a function for various values of input or model parameters. This illustration shows the response surface for a single feature f and two model parameters p_1 and p_2.

Box A
Three approaches to Map the Response Surface

Brute force sampling (not computationally feasible)

$n \approx k^p$

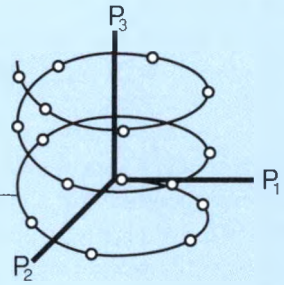
n = # of code runs
p = # of parameters
k = # of samples/parameter



Fourier Amplitude Sensitivity Test (FAST) is a global technique producing quantitative parameter sensitivities including coupling

$n \approx p^{2.6}$

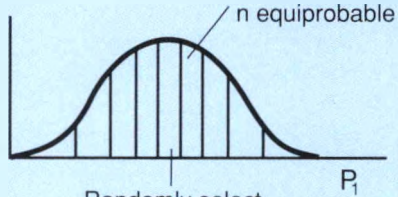
1- D search trajectory



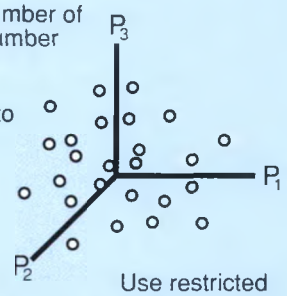
Latin Hypercube Sampling (LHS) reduces number of samples needed - allows culling of a large number of parameters

$n \approx 4/3\ p$

Divide distribution into n equiprobable cells



Randomly select value within each cell



Use restricted pairing to avoid parameter correlations

uncertain parameter may be compared to examine their relative importance.

The number of required model simulations for a FAST sensitivity analysis grows rapidly as the number of parameters of interest increases. For example, a four-parameter search requires 47 model simulations, a six-parameter search requires 99, and a ten-parameter search requires 411 model simulations. Because of these computational requirements, we must select a partial set of parameters based on prior trend analysis, physical intuition, or other sampling techniques. The number of required simulations, however, is independent of the number of output features.

We have also applied Latin Hypercube Sampling (LHS) coupled with linear fits of the response surface. LHS is a form of stratified Monte Carlo sampling, well suited for sensitivity analysis of large computer codes containing hundreds of parameters. This technique has been developed and refined in the work of R. L. Iman at Sandia and successfully applied to nuclear waste disposal and nuclear power plant risk analysis. LHS provides a basis for sparse sampling of input parameter values to map the output distribution.

Figure 6 shows a schematic of the sensitivity analysis procedure used for both techniques. First we specify the parameters for sensitivity analysis. This specification includes nominal values for each parameter and the parameter uncertainty distribution. Next, we run the model code repetitively for each set of parameters. We save the output and image files and extract the plume features from the image files. We then compute the partial variances associated with each feature using Fourier analysis techniques for FAST analyses or compute hyperplane fits for LHS analyses. These computations yield the parameter sensitivities.

Computer time requirements are principally associated with the time for the repeated code executions. The sensitivity analyses have been executed on a network of VAX 8700 class machines using Supernet. As described elsewhere in this issue (see *Supernet* pp 10-12), Supernet is a technique for linking a system of networked computers together as a "parallel processor." The sensitivity analysis problem is well suited to this type of computational resource since, in general, the plume calculations are independent of one another.

Examples of results from two formal sensitivity techniques are presented. In the first example, the Fourier Amplitude Sensitivity Test was applied to examine the coupled uncertainties of six representative parameters. In this case, the effect of dissimilar uncertainties can be compared, since we examine both macroscopic system parameters and microscopic physical parameters. In the second example, we applied Latin Hypercube Sampling to prioritize planned research on specific vibrational excitation processes.

FAST results highlight interactions among a family of system parameters.

We applied FAST to plume calculations for a representative solid-propellant rocket. Signatures of solid-fueled rockets are dominated by the emission from hot solid particles, typically alumina. Six plume code parameters were selected for evaluation with the FAST global approach: 1) missile altitude, 2) missile aspect angle, 3) the exit plane species concentration of water, 4) the exit plane concentration of carbon dioxide, 5) the mean particle size, and 6) the imaginary part of the refractive index of solid alumina. These parameters were selected to be representative of the principal types associated with the calculation of a plume signature. These types include starting conditions, motor definition, and pure physical parameters.

Computer results for this analysis indicate the influence of these uncertainties on the predicted signature of the solid rocket. The cover photograph illustrates the

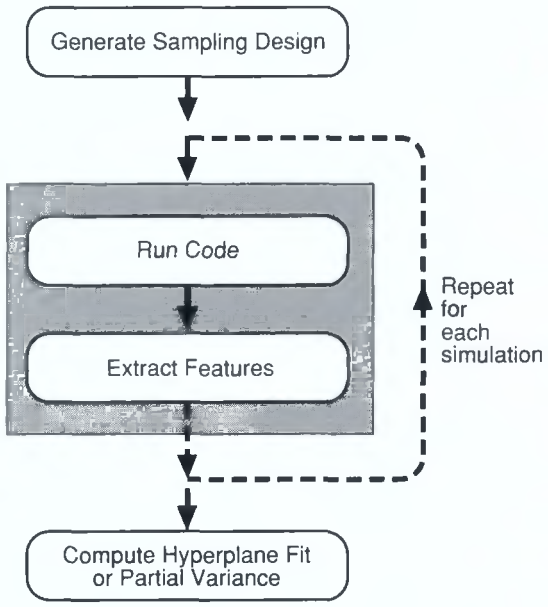


Figure 6. The procedure for sensitivity analysis of rocket plume signatures includes four principal steps.

infrared signature calculated by each simulation. The image resulting from a calculation with all parameters at their nominal conditions is enclosed in a white box. Based on the sensitivity analysis procedure, each of these calculated plumes is a possible signature for this system, yet each signature is quite different.

FAST results also may be used to assess the robustness of image features. The two distributions shown in Figure 7(A) and (B) illustrate two histograms for two different image features–the total intensity of the plume radiation (7A) and the distance from the rocket to the radiance centroid (7B). (The centroid is the location in the image of the "average" radiance value.) The results are plotted as a histogram with the value of the total intensity on the abscissa and the fraction of the total of 99 runs with a particular range of intensity value on the ordinate. The pie chart to the right of the figure indicates the partial variances due to each of the six uncertain parameters or to coupling between two of the six. The mean value μ of the feature, the standard deviation σ, and the ratio of the two μ/σ are also listed in each figure. This ratio is like a signal-to-noise ratio and can be used as a quantitative measure of the feature's robustness.

As indicated on the top right of Figure 7(A), the mean-to-variance ratio for the total intensity is 3, while in Figure 7(B), for the distance to the centroid, this same parameter is 9. Thus the centroid distance feature is more tolerant of the uncertainties in the six model parameters. These features do not provide sensor systems with the same information, so it is not a direct trade-off. But it is an indication of feature robustness that must be included in system designs.

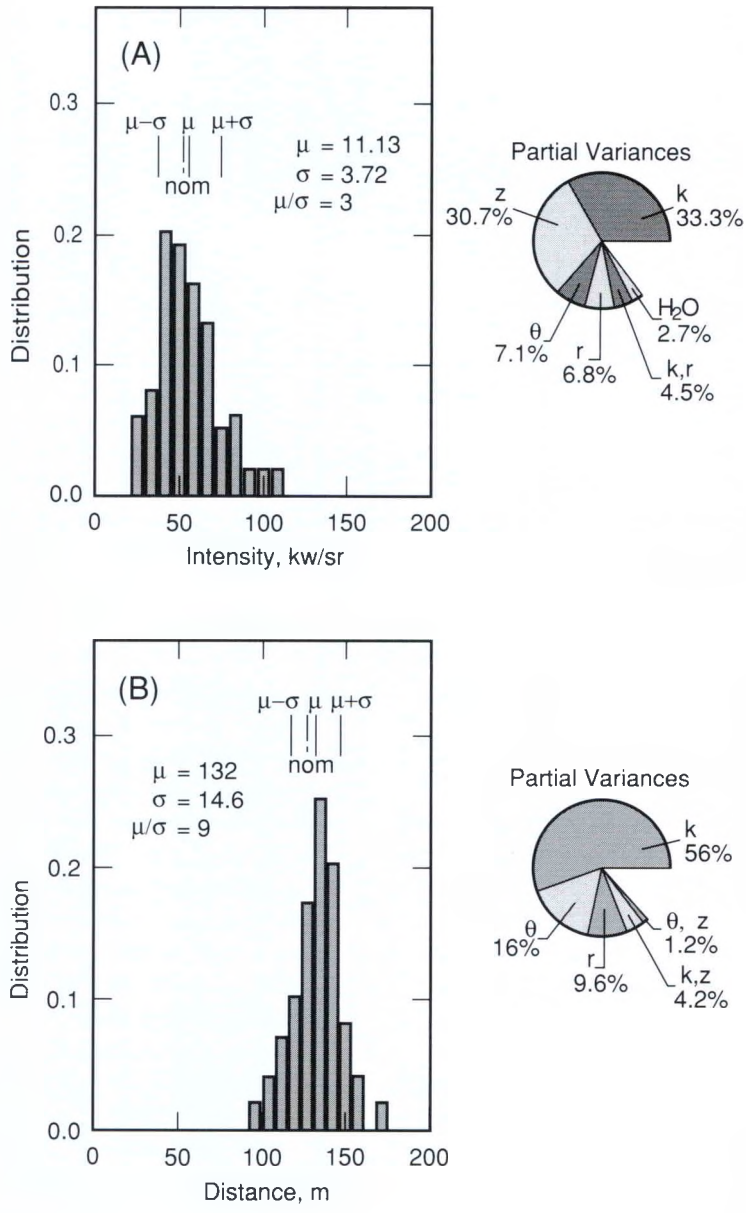


Figure 7.
These histograms illustrate sensitivity analysis results from Fourier Amplitude Sensitivity Tests applied to a solid-propellant rocket. We compared two image features–the total integrated intensity and the distance from the rocket exit to the image radiance centroid. The histograms show variations of the results for each feature, and also identify the parameters that cause the variability. The varied parameters are the missile altitude (z), the viewing aspect angle (θ), the mean particle size (r), the imaginary portion of the particle refractive index of solid alumina (k), and the exit plane concentration of water (H_20).

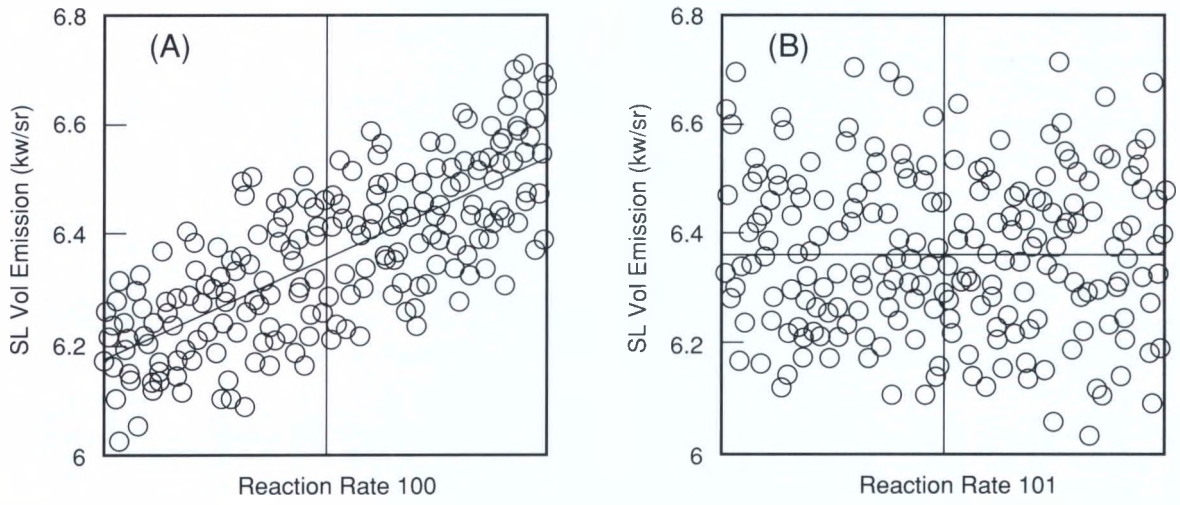


Figure 8.
These scatter plots illustrate the results from a Latin Hypercube Sampling sensitivity analysis applied to a liquid-propellant rocket. They show the total intensity in the plume as a function of two (of 179) representative vibrational excitation rates. Figure 8(A) illustrates results for a sensitive reaction rate while Figure 8(B) illustrates results for an insensitive rate.

Latin Hypercube Sampling (LHS) results are used to prioritize sensitive physical parameters.

A significant issue for program managers is how to prioritize research and development activities that are necessary to upgrade the physical models. Many organizations are working to improve the knowledge base upon which the plume codes rest. Unfortunately, the number of unknowns often outstrips the resources available at any one time for research. Thus, in light of pressing needs from the system designers for accurate code models, code development program managers desire information to allocate resources wisely in support of the hardware designers. An example of this for plumes is our imprecise knowledge of the mechanisms and rates of vibrational excitation processes which occur in the rocket plumes.

Vibrationally excited molecules of water and carbon dioxide are two of the sources of the observed radiation signatures. Currently, the CHARM code includes a quasi-two-level model of vibrational energy which is used to obtain excited-state populations for radiation calculations. The steady-state populations of each ground and excited state are computed through reference to a tabulation of temperature-dependent reaction rates for 179 exchange paths for vibrational energy transfer. In most cases, the rates, and their temperature dependence in particular, are not well known despite laboratory studies and *first principle* calculations. Uncertainties can be an order of magnitude or greater. Additional laboratory and analytical research is to be performed, but the question is which reaction processes should be emphasized.

To answer this question with quantitative information, we examined reaction rate sensitivities for a large, liquid-fueled rocket at conditions leading to substantial enhancement region radiation. All 179 vibrational excitation rates were selected as uncertain parameters; 240 runs of the model were executed. We examined output features in five spectral bands.

We analyzed the computed plume data through a multi-dimensional linear fit of the feature values with parameters as independent variables using a linear least squares regression analysis. Separate hyperplanes were fitted for each of the spectral bands.

Although the sensitivity analysis is automated, visualization of the intermediate results is both informative and an important verification of proper code performance. Figure 8 shows two selected scatter plots for the 240 runs described above. There is one scatter plot for each parameter, or 179 plots in this case. The ordinate in each plot is the computed feature value. The abscissas are the parameter values arranged in ascending order. One is looking for a qualitative trend in the data. Clearly, Figure 8(A) shows such a trend while Figure 8(B) does not. The parameter represented in Figure 8(A), the reaction rate of process 100, has a high sensitivity while the parameter represented in Figure (8B) has a low sensitivity. The scatter plots also show the computed hyperplane fit.

Quantitative results from this analysis highlight five key reactions involving the rates of water and carbon dioxide to which the intensity is quite sensitive. Sensitivities as large as 0.3 were computed. This means that if the rate of a particular reaction, for example, is uncertain by an order of magnitude, then the predicted plume intensity will be uncertain by a factor of three from that one reaction rate alone (aside from uncertainties in all other parameters). The specific results from this study provide quantitative information on those rates whose values must be precisely known.

Sensitivity results are used by both system designers and researchers.

An operational Strategic Defense System must perform robustly in light of uncertainties arising from imprecise knowledge of threat systems, environmental variabilities and our lack of full understanding of the physical processes. Our sensitivity analysis results contribute directly to system reliability. As illustrated above, they are being used by experiment designers, SDI system designers, and scientists performing supporting research. An additional advantage of the use of sensitivity techniques is improved reliability of the plume codes themselves since they are tested over a large envelope of conditions. Prior to this work, codes were often exercised about only a few point design conditions.

In related work, we are using the sensitivity results coupled with assessments of proposed targeting algorithms to ensure that relatively invariant features in the plume images are used. The payoff for the strategic defense community is greater confidence in algorithms that will be robust in the face of uncertainties.

For further information, please call ***Patricia K. Falcone,*** (415) 294-2524.

SUPERNET: Supercomputer Performance with Parallel Processing on a Network of Minicomputers

Sandia's parallel processing manager, called "Supernet," utilizes 14 minicomputers on a network that spans 1100 miles to provide supercomputer capability.

Computer scientists at Sandia National Laboratories have demonstrated a network parallel processing manager termed "Supernet" that achieves supercomputer performance for some major applications problems. This demonstration used fourteen DEC VAX minicomputers on a network that spans 1,100 miles over a 56 kilobit link (see Figure 1). Supernet provides adjunct supercomputer capability by using existing network minicomputers around the clock. This concept is of interest because supercomputers are very expensive resources costing upwards of $20 million per system. Harnessing an organization's existing minicomputers for parallel computations may prove to be a cost-effective supplement to supercomputers.

The managing software controlling execution of parallel segments on "slave" computers from a "master" machine is the VAX VMS version of "Linda-C," developed by David Gelernter, Nicholas Carriero, and Jerry Leichter of Yale University. In an article entitled "Programming for Advanced Computing," SCIENTIFIC AMERICAN, October 1987, Gelernter describes the Linda "tuple space" model in detail. The recent Sandia success is an encouraging demonstration of its feasibility for large applications.

Linda provides a simple elegant method of expressing the way in which the various portions of a parallel program communicate with each other. The programmer segments an application into several pieces that can be performed in parallel.

The program accomplishes communication by manipulating tuples and tuple space. A tuple is an ordered list of data fields. The data fields contained in a tuple may be character strings, integers, or floating point representations (similar to scientific notation) for computations. Passive tuples hold the common data structures of the executing system, while active tuples act as message passers by creating or consuming passive tuples. "Tuple space" is shared memory that can be

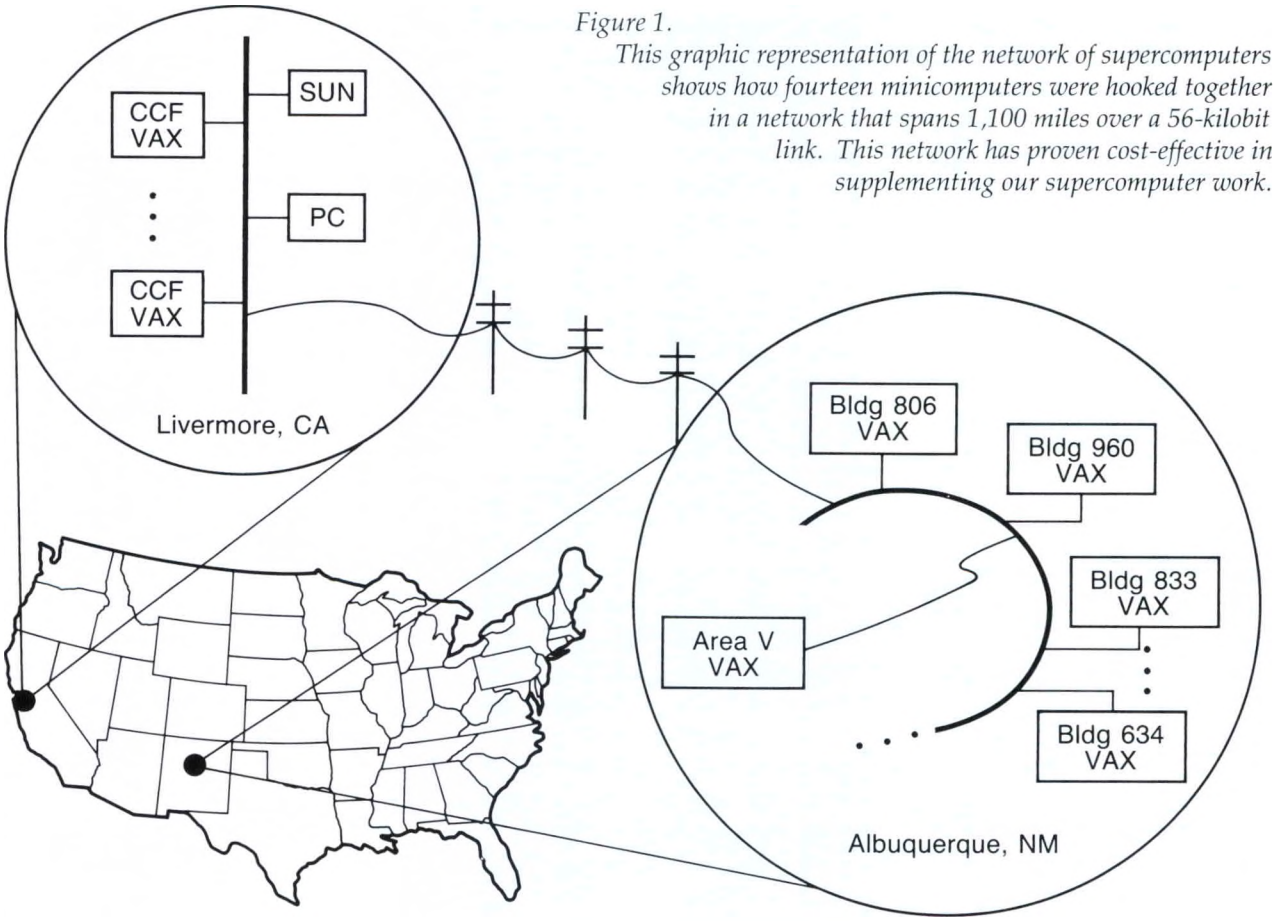


Figure 1.
This graphic representation of the network of supercomputers shows how fourteen minicomputers were hooked together in a network that spans 1,100 miles over a 56-kilobit link. This network has proven cost-effective in supplementing our supercomputer work.

thought of (in total) as a bulletin board where tuples are posted by executing segments of code. Tuple space may be distributed among several processors.

Linda is not a stand-alone programming language; rather, it is a set of operators that are injected into an existing language such as C or Fortran. Tuples are inserted into tuple space with the "out" operator, read with "rd," and extracted from tuple space with "in." The tuple definition specified in an "rd" or "in" command serves as a template to which tuples in tuple space are matched.

Sandia used Supernet to benchmark two applications, BOHR and CHARM, that currently run on a Cray-1S supercomputer. These codes divide easily into large, independent pieces and give encouraging performance results, as shown in the table.

We use BOHR at Sandia to study early events in radiation damage of semiconductors. These computations examine the interaction between a high-energy incoming projectile and a target system. For example, in one computation, the projectile was an argon nucleus and the target was a silicon atom. This Monte Carlo code uses a few random numbers to initialize the projectile direction and velocity. It then follows the trajectory by numerically integrating the classical equations of motion until the collision is completed. We then assess results of the collision: electrons in the target, for instance, may have been ejected, excited, or transferred to the projectile, generating recoil energy in the target which is then computed. After this, another trajectory is set up and followed. One run of BOHR consists of following thousands of trajectories for which various possible outcome results are averaged, counted, and tabulated.

The parallel implementation of BOHR is straightforward because each trajectory is completely independent of the next. We employ

Application	Cray-1 Execution Time (sec)	Supernet Execution Time (sec)	Cray Equivalent
BOHR	7.4	15.0	0.5
CHARM	343.0	143.0	2.4

a single master process and many worker processes–one on each machine that participates in the computation. Each slave process repeatedly follows one trajectory, and sends the results back to the master process. When the master process has received and tabulated the required number of trajectory results, it writes the output file.

One useful aspect of this implementation is its robustness in the face of inevitable difficulties with machines on the network. Even though a machine running a slave process crashes, the computation as a whole proceeds. Although the master process receives no results from that one slave, eventually enough trajectories are received from other machines to complete the job.

The BOHR program vectorizes well on a Cray supercomputer, which runs the code about twenty times faster than a DEC VAX 8700 minicomputer. A Supernet set of four DEC VAX 8700s and seven 8650s achieved half the speed of the Cray for this problem.

CHARM is a computer program for modeling high altitude rocket plumes (see the article beginning on Page 2). The model has a number of parameters, and a study underway at Sandia seeks to determine the sensitivity of plume simulation results to the values of these parameters. The computationally intensive portion of the study reduces to running CHARM hundreds of times with different input values. Because this particular application vectorizes poorly on a Cray-1S, we were able to achieve an improvement of 2.4 times on Supernet.

The parallel implementation is once again straightforward. Three types of processes are employed: one master, one post-processor, and one slave process on each worker VAX. Each slave participating in the computation posts a Linda tuple indicating that it is idle; then waits for a response from the master. The master process waits for such "Idle VAX" tuples to appear and sends each such process the index for the next simulation run to be performed. When the run is completed, the slave posts a message to the post-processor indicating that the output from the CHARM run is available for analysis. It then requests additional work from the master.

This application also is robust in the face of difficulties with a remote slave processor. The master assigns simulation runs to slaves until all have been assigned. At this point, although all of the runs have been assigned, not all of them have completed. When a slave process requests more work, it is assigned one of the not-yet-completed simulations. If one slave never completes its task, it will simply not report back to the master, but its task will be reassigned to another processor, and the job as a whole will complete.

The table shows execution time for a sensitivity analysis requiring 99 runs of the CHARM program. The Supernet set of fourteen 8000-series DEC VAXes performed the problem more than twice as fast as the Cray. For this application, the network delivered true supercomputer performance.

A benchmark of a tightly coupled, finite-element, thermal analysis code (TACO3D) developed for execution on a Hypercube distributed-memory parallel processor gave poor results, disclosing Supernet limitations. Supernet processors are fast, but network communications are very slow. Applications best suited to run on Supernet are those that can be divided into large, computationally intensive segments that are not tightly coupled to the results of other segments.

Even though the efficiency of Supernet varies with the application, the cost-effectiveness gained by utilizing the idle cycles on an existing network of distributed computers makes it a worthwhile supplement to supercomputers. Supernet is potentially expandable to a large set of heterogeneous computers such as personal computers and engineering work stations. The original implementation of Linda-C was limited to a single physical Ethernet, since it was felt that the performance using a direct Ethernet connection would be about as slow as anyone would find useful. However, it has become clear that some applications can utilize a large number of machines even with very high communications costs. We have, therefore, used DECnet, task-to-task communication to extend a subset of Linda operations to more remote VMS machines and TCP/IP tools to access UNIX machines. A high priority item for future work is to enhance these extensions to support all the Linda operations and merge them into the base system.

For more information, please call *W. D. Wilson,* (415) 294-2264.

Starloc: A Special-Purpose Computer for Automatic Target Recognition

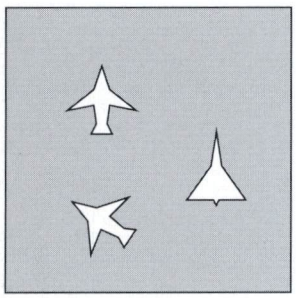
We are developing Starloc—a Sandia Target Location Computer that swiftly locates targets in an input image.

Automatic target recognition (ATR) has a wide range of applications, from advanced military concepts for autonomous (unmanned) vehicles to robotics and medical imaging. ATR uses a television camera or other imaging device to record a two-dimensional image of a (usually) three-dimensional scene. This input image consists of some number of objects and the background. The general goal of automatic target recognition is to identify the objects in the input image. However, a particularly important part of the problem is to determine whether or not a certain type of object is in the image and the quantity of these objects. For example, if the input scene is a battleground, then the application of ATR to the input image can identify how many tanks there are in that scene.

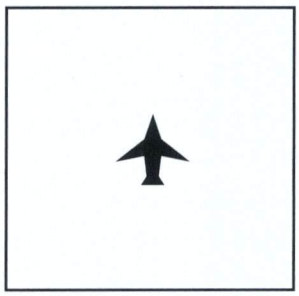
There are many proposed methods for automatic target recognition. One commonly used method is called template matching, illustrated in Figure 1. The template is an image copy of the object being sought. When the template is laid over an image object for which it was designed, we get a template match. The process of checking the template against all possible positions in the image is called the correlation of the input image and the template. Using correlation for each position of the template, we obtain a measure of the match between the template and what it lies over. The measure is called the correlation value, and the set of correlation values for all positions of the template is the correlation result.

The template matching method encounters problems when there are variations in the object's image. When the object is rotated or changed in scale (size), the object's image no longer matches the template. One possible solution to this problem is to use multiple templates, one for each variation. However, the number of templates required increases geometrically as the number of variation types to be accommodated grows. For example, if 360 templates are sufficient to match any rotation and 100 templates are sufficient to match any scale, then 100 x 360 or 36,000 templates are required to match any rotation at any scale. Each one of these templates has to be correlated with the input image. When other variations such as the object brightness and the angle-of-view of the object are considered, this approach requires a very large number of correlations.

In previous work, Sandia developed a generalization of the template matching algorithm called the Lock-and-Tumbler (LAT) filter algorithm. The LAT algorithm mathematically combines the thousands of templates needed to recognize a given object subjected to a given set of

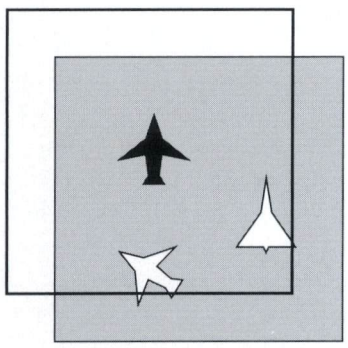


Image

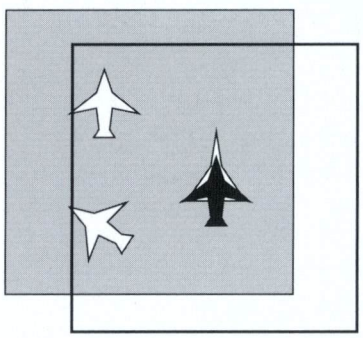


Template

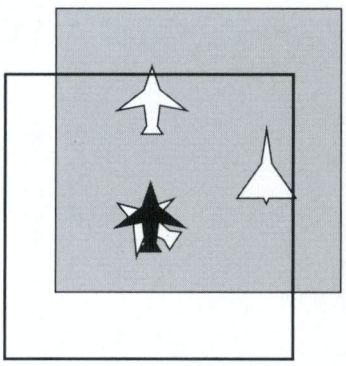
Figure 1.
The simple template matching algorithm passes a template designed for a specific object over an image. When the template is directly over the object, the match is good. When the template is over a different object, or when the template is over a rotated version of its own object, the match is poor.



Good Template Match with Image



Poor Template Match with Image (wrong object)



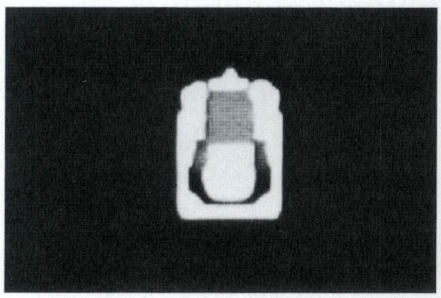
Poor Template Match with Image (rotated object)

image variations such as rotation, scale, brightness (intensity), and angle-of-view into a small set (ten to twenty) of complex templates called filter kernels. An example target (a bulldozer) and one of its complex LAT filter kernels are shown in Figure 2. We then correlate this set of filter kernels with the input image. Objects are found when the correlation values of the object's image with each of the filter kernels are within acceptable limits. The desired object "keys in" with the filter kernels just as the notches in a key "key in" with the tumblers in a lock.

Motivated by the power of the LAT algorithms, we are building a special-purpose computer to perform this generalized template matching algorithm. Each correlation of the input image with a complex filter kernel requires millions of mathematical operations. For many applications, a general purpose computer cannot perform the operations fast enough to be useful. The computer we are developing is called Starloc (the Sandia Target Location Computer), and it achieves excellent speed performance using only readily available, commercial integrated circuits. We have constructed a reduced-scale prototype that can perform 40 million floating-point operations per second. In addition, we have included features to make the computer fault-tolerant so that it will perform robustly in the field. These fault-tolerant features can detect computational errors and reconfigure the computer around damaged components.

Starloc's algorithmic design recognizes a target in an image subjected to a specified set of variations.

To correlate the input image with the set of filter kernels computationally, the input image is sampled at discrete locations, typically on a square n pixel by n pixel grid. A typical value for n is 256, and the input image can thus be represented by a 256 x 256 array of numbers where the numbers correspond to the brightness (intensity)



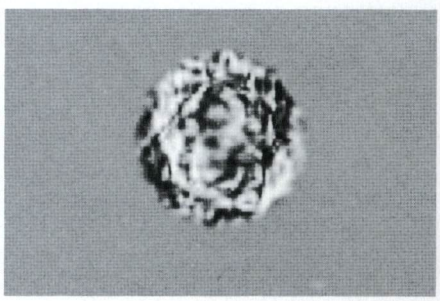


Figure 2.
The Lock-and-Tumbler filter algorithm uses a set of complex filter kernel templates to correlate with an input image. Above is a desired target (a bulldozer) and the real part of one of its complex filter kernels.

of the image at the selected pixels. The input image is composed of targets (desired objects) and background. In this more restrictive definition, the background includes all objects that are not being sought. Examples of targets on backgrounds include a rocket plume in space, a tank on a battlefield, an abstract character printed on a page of paper, or a part traveling along a conveyor belt on an automated assembly line.

The LAT filter algorithm is implemented in four modules as shown in Figure 3. In the Image Reception Module, the input scene is imaged onto a sensor where the image is discretized into pixels and the analog sensor signals at each pixel are converted to a digital representation. The image is enhanced in the Digital Enhancement Module. The type of enhancement depends on the type of input scene and features used to identify targets. Typical enhancements include filtering for noise removal and edge enhancement. The Correlation Module performs the correlation of the enhanced image with the filter kernels. Targets are located at pixels that display similar correlation values for identical positions of all filters in the filter set. The correlation value similarity is determined by calculating the ratio of the value mean to the value standard deviation for each output pixel location. Where the actual pixel values are nearly constant across all of the kernels, the standard deviation is small, and thus the ratio is large.

The mean and standard deviation of a series of numbers can be calculated from the sum of the numbers, the sum of the squares of the numbers, and the number of numbers in the series. The Correlation Module computes the sum of the correlation values and the sum of the squares of the correlation values for each output pixel location over the filter set. The Digital Post-processing Module produces the

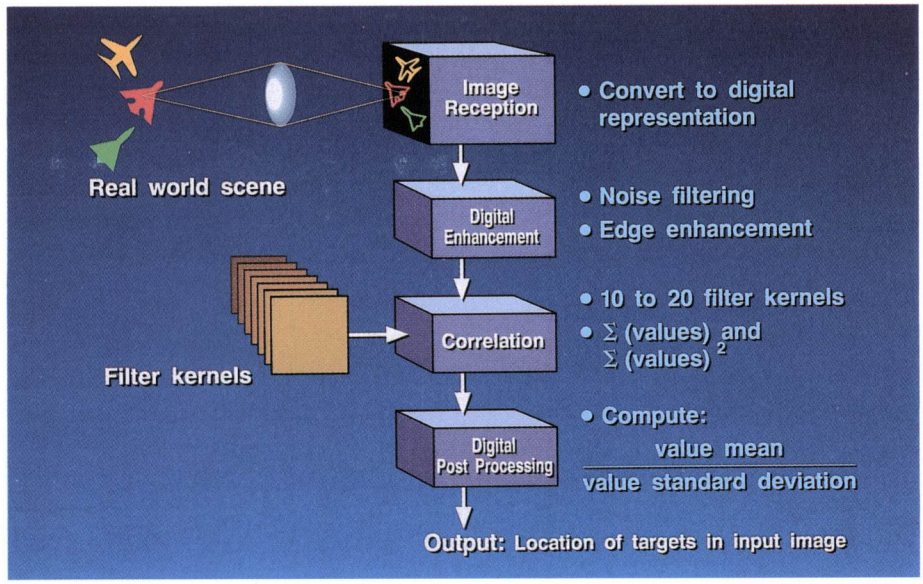


Figure 3.
The Lock-and-Tumbler filter algorithm is performed in four modules. Of these, the Correlation Module is the most computationally demanding.

LAT filter output. From the running sums and the number of kernels in the filter set, the module calculates the ratio of the value mean to the value standard deviation for each pixel location.

Starloc performs the operations of the Correlation Module, the most computationally demanding module. A correlation can be performed in either of two ways: 1) directly or 2) in the Fourier domain by first discrete Fourier transforming the image and filter kernel into the Fourier domain, multiplying the transformed image and kernel pixel location by pixel location, and finally inverse transforming the product. For increased efficiency, the forward and inverse discrete Fourier transforms are performed using Fast Fourier Transform (FFT) algorithms. We chose the Fourier domain method because it requires fewer operations for the image and target sizes in which we are interested.

The basic functional blocks required to digitally perform the LAT filter Correlation Module are shown in Figure 4. Using FFT, the image is transformed into the Fourier domain. The transformed image and a previously transformed filter kernel are then multiplied pixel by pixel, and the inverse product is computed using FFT. For the LAT filter algorithm, a set of kernels is correlated with the input image, and the sum of correlation values and correlation values squared is computed and stored. The description of the algorithm has only four functional blocks: the FFT of the input image, the pixel by pixel multiplication, the inverse FFT, and the collection of results. Even when using a computationally efficient FFT algorithm, the computational demands are quite large. Approximately 100 million floating-point operations are required for each 256 x 256 pixel image when using an 18-filter, precomputed filter set.

Starloc's architecture promises to meet our objectives for high performance, efficiency, and fault tolerance.

The Starloc architectural design attempts to achieve three objectives: high performance, efficiency, and fault tolerance. High performance is dictated by the considerable arithmetic processing requirements of the LAT algorithm. No single processor could perform all of the operations in any "real time" sense. The architecture, therefore, must combine the efforts of many processors in an organized fashion, distributing the algorithmic work load evenly across the processors. Idle processors represent wasted computational capacity and thus computational inefficiency. Fault tolerance implies two architectural features: the ability to detect faults and the ability to reconfigure the architecture to remove the effects of the faulty units.

Starloc's fault tolerance philosophy arises from one specific goal: the continued operation of Starloc at its rated performance in the presence of any single faulty unit. (When possible, continued operation, albeit with possible performance degradation, is also desired for multiple concurrent faults.) To achieve maximum performance and efficiency, fault tolerance techniques based on time redundancy (performing all the computations multiple times) or massive hardware redundancy are not appropriate. The fault tolerance method in Starloc uses some hardware redundancy, but attempts to do so with a minimum amount of hardware overhead.

Fault checkers constitute part of the hardware overhead needed for fault detection. One fundamental problem in fault detection is "Who checks the checkers?" To avoid multiple checkers for each unit, the hardware and software are designed so that each checker checks several possible faults–its "fault set." The sets are constructed so that any single fault is a member of more than one fault set, and, therefore, will be detected by more than one checker. A fault detection signal produced by only one checker implies that one of the checkers has failed. When a fault is detected, there are several options available. The selected option is determined by the nature of the fault and its position in the processing chain. If the fault is permanent, then reconfiguration of the data and control paths through the processors is necessary to route the paths around the damaged area. Hardware bypasses and standby processing stages are included in the design to facilitate the reconfiguration.

Figure 4 shows a diagram of the LAT filter algorithm from which the general architecture of Starloc is derived. The algorithmic stages form a linear, branchless, data processing path from image input to correlated output. The basic data-set being operated upon within each stage is a 256 x 256 pixel image frame. In Starloc, each algorithmic stage is assigned to a processor stage. There are ten processor stages: two pixel by pixel (PxP) processor stages and eight FFT processor stages. One PxP stage performs the multiplication of the input image FFT and the filter kernel FFT. The other PxP stage sums the correlation values and values squared at each pixel location over the set of filters. Because of the similarity between the forward and inverse FFTs, two FFT algorithmic stages, one forward and one inverse, are assigned to each FFT processing stage.

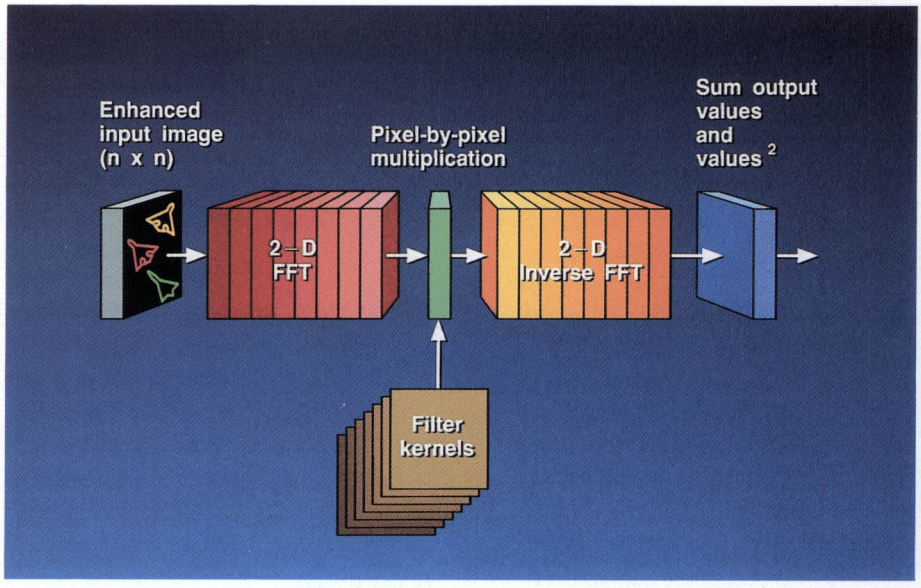


Figure 4.
Starloc performs the correlation module operations. The correlation is done by transforming the input image into the Fourier domain, multiplying it by a filter kernel already in the Fourier domain, and then inverse transforming the product back to the space domain. The search for correlation value constancy requires collection of output statistics pixel by pixel. The mean and standard deviation of each output pixel value are obtained from the sum of the values and the sum of the squares of the values.

The general topology of the architecture is a circularly connected ring of processor stages with some branches into and out of the ring. Figure 5 shows the flow of information through the processor stages. Movement of data around the ring is in one direction only. We reduced the input/output interface complexity of Starloc by making the input image actually enter the computer through the PxP multiplication stage through which it passes with no processing taking place. The data is pipelined through the computer. Each stage receives and temporarily stores a data frame from the previous stage, performs some operation on the data frame a few elements at a time, and sends the results on to the next stage. The effect is like a computational assembly line where the line inputs are the input images, the line outputs are the target locations, and the assembly line stations are the processing stages.

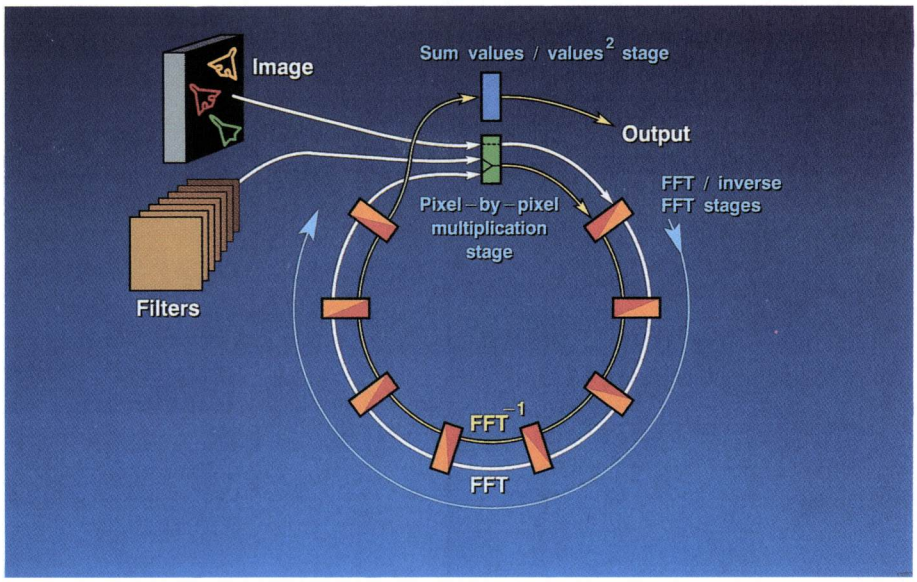


Figure 5.
Starloc's architecture is a circularly connected set of processing stages. Data flows through the processing stages and is operated upon in an assembly line fashion. At any one time, each processing stage is operating on a different image/kernel correlation.

We simplified the design of Starloc by making the basic structure of all stages identical. Differences between FFT stages and PxP stages are restricted to a small number of the components within each stage. This design similarity between stages does not adversely affect architectural efficiency since each stage processes an entire image-sized frame and each stage has about the same amount of processing work to do. Figure 6 shows the six major subsystems in each stage. These subsystems are the input selector, the controller, the memory, the address generator, the register files/floating-point processors, and the first-in-first-out (FIFO) queues. Control information and data move among the subsystems as indicated in the figure.

The stage input selector is a multiplexer for input data and control signals. The input data consist of either an input image, a filter kernel, or the results computed by the previous processing stage. The control signals include handoff control signals for data moving between stages, addresses for data writes in the current stage memory for data produced by the previous processing stage, and control information about the incoming data frame.

The controller provides the control signals for all of the other subsystems within a stage, either directly or indirectly. The controller in each stage is a fully self-contained, high-speed microcontroller consisting of a sequencer and a control store, and each stage can have a different set of control instructions. The control unit is horizontally microprogrammed. The long control word (184 bits wide) contains positions for all control lines to the address generator, registers, and floating-point processors.

The memory in each stage stores a complete image or image-filter product in the 64-bit complex floating-point format consisting of a 32-bit real number and a 32-bit imaginary number. Since the memory receives results from the previous stage while it is supplying data for processing in its own stage, there is a strong possibility for memory conflicts (i.e., having the previous stage write over a value in memory before it is used in the current stage). The memory is interleaved to alleviate this problem. The memory subsystem is actually two memories, called memory A and memory B, that share the same address space. One memory is used for the destination of products from the previous stage and the other as the source of operands for the current stage. As data propagates down the pipeline, the A and B memories alternate between source and destination memories; after a destination memory becomes filled with the results from one stage, it becomes the source of data for the next stage.

The address generator produces all of the addresses for data entering or exiting the registers/floating-point processors in a particular stage. The generator thus produces the read addresses for the memory in the current stage and the write addresses for the memory in the subsequent stage.

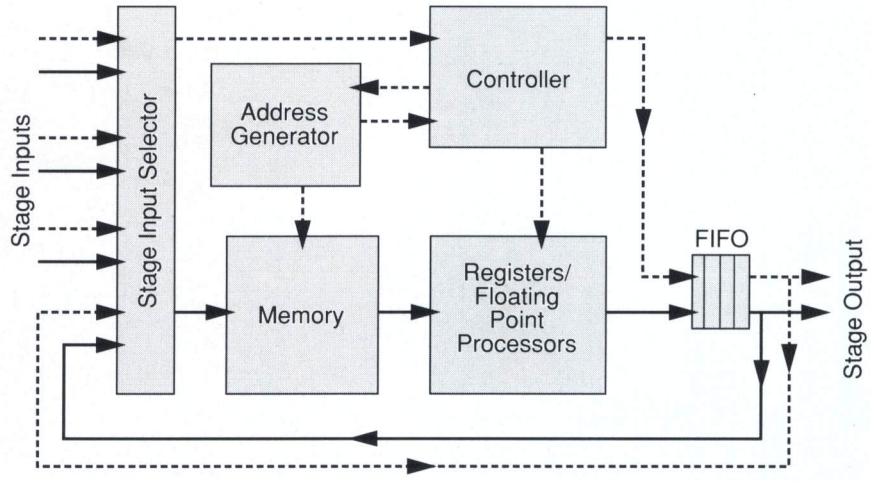


Figure 6.
The basic structure of each processing stage is the same. Each stage is a self-contained computer with input selection and queued output, memory and memory address generation, control, and high-speed data processing. In this figure, the flow of data is shown by the solid lines, and the flow of control is shown by the dashed lines.

The registers/floating-point processors module consists of a five-port register file, a complex switching network called the data router, and two MFLOP (million floating-point operations per second) processors. In one major cycle, the register file can 1) receive one complex word from the stage's memory, 2) send one complex word to the next stage's memory, 3) and 4) send two complex words to the floating-point processors, and 5) receive one complex word from the floating-point processors. Between the register file and the floating-point processors, the complex words are broken into their 32-bit real and imaginary components. Each of these components can be routed to any of the inputs of the floating-point processors through the data router. The two floating-point processors can each perform one floating-point addition, subtraction, or multiplication in 100 ns.

The output for each stage is routed to the two subsequent stages and to itself. (The second subsequent stage path is used if the first subsequent stage is faulty.) The FIFO queues allow the output of each stage to be somewhat decoupled from the input of the subsequent stage. The queues act as a buffer so that output demands of one stage do not have to be met instantaneously by the input capacity of the subsequent stage.

We have designed and constructed a prototype system to assess the impact of fault-tolerant design overhead and provide hardware support.

A prototype system that has full functionality but fewer stages has been designed and constructed (Figure 7). The system contains two processing stages, one devoted to the FFT and the other to the PxP stages, although both are very similar. Each is identical with the stage design just described and contains all fault-detection features. However, the prototype has lower performance and if faults are detected cannot architecturally reconfigure the system. The prototype system also includes an interface designed for a DEC MicroVAX host. The connections between stages and between stages and the host are shown in Figure 8.

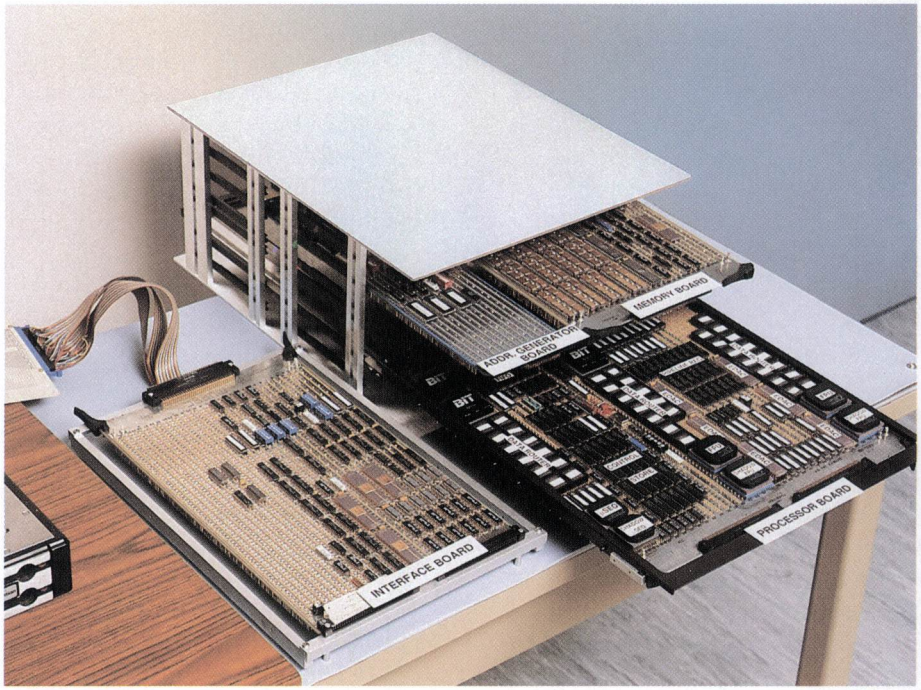


Figure 7.
The 40 MFLOP prototype has only two stages in a single chassis. This photograph shows the chassis with the board set for one processing stage and a separate board to interface with a host computer.

We have several goals for the prototype. One is to assess the impact of fault-tolerant design overhead. Another is to provide hardware support for further ATR algorithm development. The prototype uses the equivalent of four full-size, wirewrap boards and consumes about 400 watts of power at a rated performance of 40 MFLOPs. Figure 9 shows a typical result when we use the LAT filter algorithm. We used a filter set designed to find bulldozers on an input scene consisting of bulldozers, a front-end loader, and trees. The red spots show identified and located target bulldozers.

Starloc mixes practical engineering guidelines with fault-tolerance principles in a complete design.

A computer that automatically recognizes targets must be fast, reliable, and compact. We have combined algorithms, architecture, and fault tolerance to satisfy these requirements in the design of Starloc. Starloc mixes practical engineering guidelines with serious fault-tolerance principles in its complete design. Although containing some special features specifically oriented toward the Lock-and-Tumbler filter algorithm, the design has a wide range of applicability to numerous image processing algorithms, particularly those based on two-dimensional correlation or two-dimensional Fast Fourier Transforms. We have designed and constructed a two-stage prototype that is supporting further ATR algorithm development while serving as a test bed for measuring fault-tolerance coverage.

For more information, please call ***Leonard M. Napolitano, Jr.,*** (415) 294-3218.

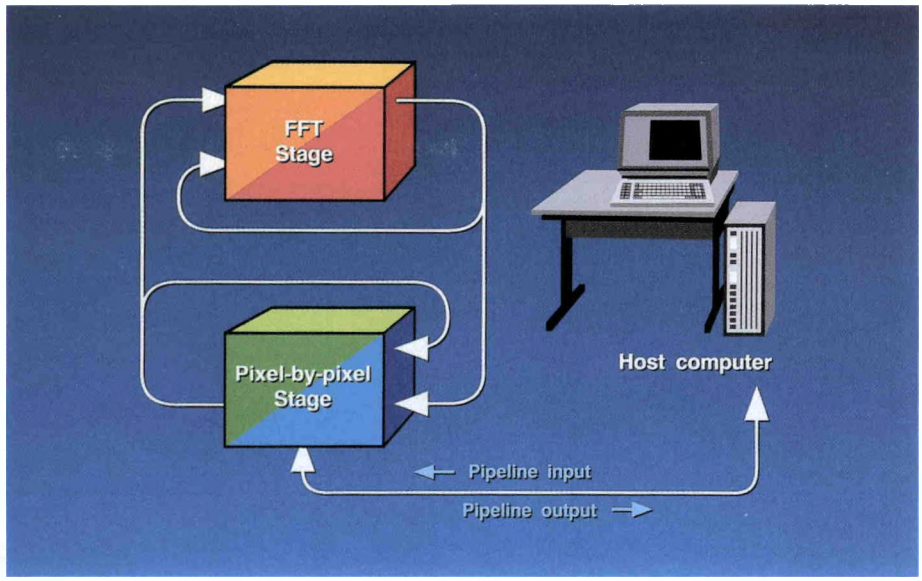


Figure 8.
The flow of data between the two stages in the prototype mimics the flow of data through all the stages in a full-up system. The prototype lacks the reconfiguration capabilities required for full fault-tolerance.

Figure 9.
This photograph shows a sample result for the Lock-and-Tumbler filter system using the Starloc prototype. The input scene consists of bulldozers, a front-end loader, and trees on an earth- and grass-colored background. The red spots designate found targets (the bulldozers).

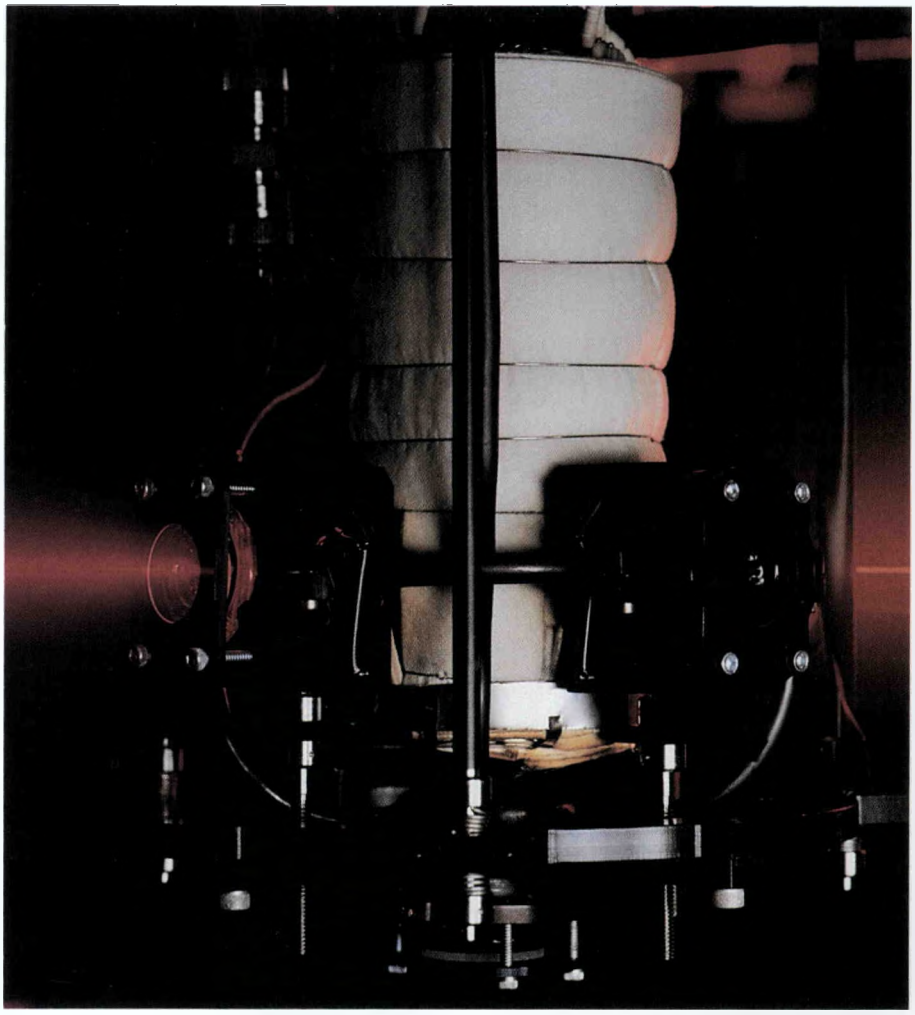


Figure 1.
This cell was developed to study the reactions of OH with fuels over a wide temperature range. Reagents and diluents enter the cell through jets that ensure good mixing. The beam from an excimer laser enters the cell through the right of the two visible ports and initiates chemistry that is probed by an ultraviolet beam following the same path as that of the red beam used for visualization.

Combustion Reactions of OH

The detailed information we are now obtaining about fuel-consuming reactions of the hydroxyl radical is helping us understand how fuels burn and why engines sometimes knock.

Over 90% of the nation's energy is derived from combustion. At Sandia's Combustion Research Facility (CRF) in Livermore, the Department of Energy (DOE) supports research that is developing the science and technology required to maximize the benefits and minimize the problems arising from the burning of fuels. Research at the CRF spans a wide variety of topics, ranging from fundamental studies of the dynamics of molecular collisions to applied work on the burning of coal.

Understanding the chemical reactions in processes that burn fuels is vital in our work.

As we face increased demands for energy and growing concerns about combustion-related pollutants in the environment, it is becoming more important to understand in detail the chemical reactions governing the efficiency of processes that burn fuels and the extent to which undesirable by-products form and can be destroyed (see Figure 1). Recent work at the CRF addressed

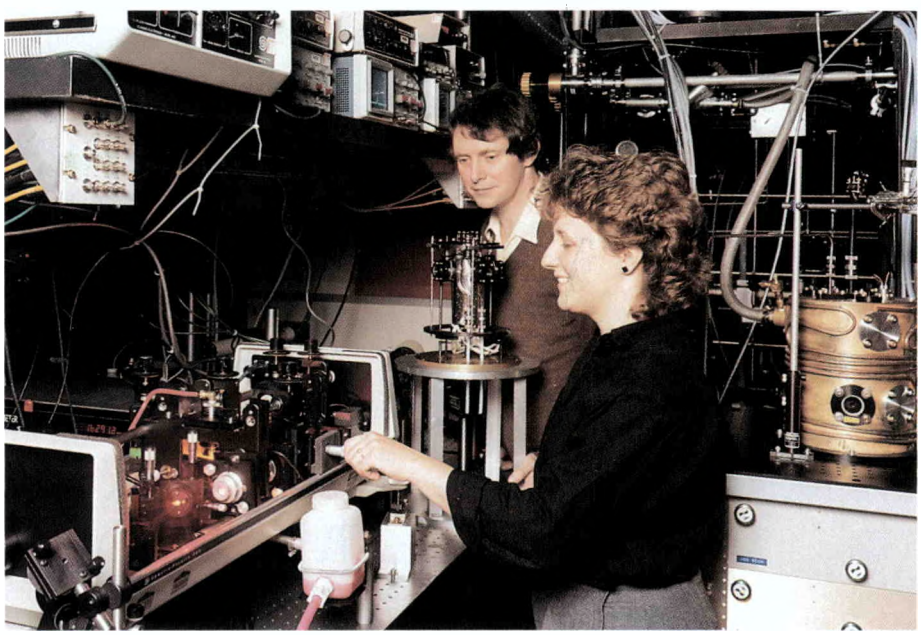
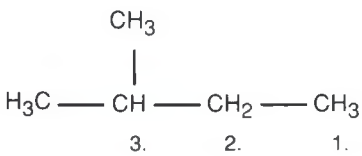
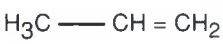


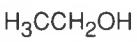
Figure 3.
Frank Tully and August Droege, a recent postdoctoral associate, make adjustments to the probe laser used in kinetics experiments on OH radicals. The reaction cell is enclosed in the brass cylinder behind Dr. Droege.



alkane: contains one single bond
1. primary C-H site; C bonds to one another C-atom
2. secondary C-H site; C bonds to two other C-atoms
3. tertiary C-H site; C bonds to three other C-atoms



alkene: contains one double bond



alcohol

Figure 2.
Alkanes contain only hydrogen and carbon with no double bonds between carbons. There are three distinct kinds of C-H bond, depending on the number of other C-atoms to which the carbon in question is bonded. Alkenes contain only C and H, but they have a double bond between carbons. Alcohols contain oxygen as well as C and H, with the O incorporated, as shown.

reactions involving nitrogen and led to the development of RAPRENOx, a process for removing NOx from exhaust gases. One effort, presently supported by DOE's Office of Basic Energy Sciences, Division of Chemical Sciences, focuses on a class of reactions that play a major role in the consumption of fuel molecules in flames. Studies of the rates of these reactions are advancing our fundamental understanding of microscopic molecular chemical processes, and the results are being used in sophisticated models of preignition knock in engines.

The hydroxyl radical, OH, is a critical and ubiquitous reactive species in combustion. In hydrocarbon flames, for example, reactions with OH usually dominate the consumption of fuel. The manner in which OH reacts with a given fuel molecule depends on the detailed molecular structure of the fuel (see Figure 2). Different pathways exist for OH reactions with alkanes, the major components of gasoline, and for alcohols, presently being used as smog-reducing additives to gasoline. Alkenes are molecules that form following an OH attack on alkanes and react with OH in still a different way.

We use laser techniques that precisely characterize the chemical kinetics of OH reactions.

At the CRF, we utilize lasers to determine the rates of reactions between OH and hydrocarbons as well as the fraction of a reaction that yields each of the possible products (see Figure 3). In our laser photolysis/laser-induced fluorescence (LP-LIF) technique (see Box A), one laser initiates the desired chemical reaction, and a second laser monitors its progress in time. The selectivity and high sensitivity of laser techniques precisely characterize the chemical kinetics of these important reactions. The rates of chemical reactions are very sensitive to temperature, and when we study a reaction over a wide temperature range, we learn a great deal about how it proceeds.

Box A
Laser Photolysis/Laser-Induced Fluorescence Technique

Developing improved diagnostic techniques is an important component of our research. Laser-based measurements are especially useful because they are versatile and can be performed without perturbing the phenomena being studied. We use lasers in kinetics experiments both to make reactive species and to determine their rates of reaction. A pulsed laser emitting 193-nm radiation initiates the chemical reaction by breaking chemical bonds and producing OH radicals. The OH radicals react with added hydrocarbon (RH) and the OH disappearance rate is monitored by a unique excitation/fluorescence method developed at Sandia that utilizes an ultraviolet, continuous-wave laser probe. The laser techniques permit each OH + RH reaction to be studied individually, and they enable us to reach levels of accuracy and chemical detail previously unattainable.

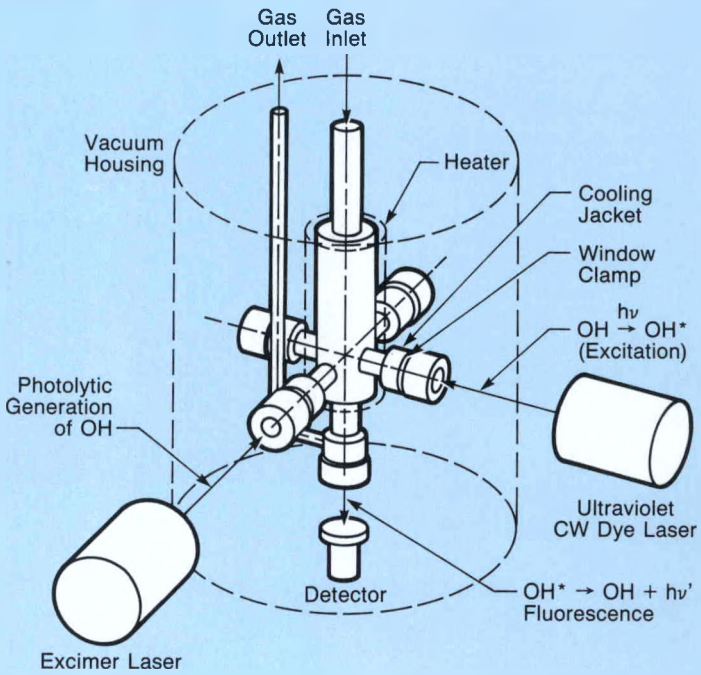


Figure A-1.
This schematic diagram shows the experimental technique. The output from the excimer laser initiates reactions in the cell shown in Figure 1. The extent of reaction is monitored using a second laser.

The simplest hydroxyl-radical reaction is hydrogen-atom abstraction from a hydrocarbon, RH, and is written OH + RH→ H_2O + P. For alkanes, which contain no multiple bonds or easily accessible electrons, hydrogen-atom abstraction by OH is the only mode of reaction. Using the LP-LIF technique, we measure the total rate of hydrogen-atom abstraction by OH from all reactive sites in the alkane. By substituting deuterium for hydrogen at selective reactive sites, we reduce reactivity at those sites, and we learn about the individual product yields resulting from abstraction from primary, secondary, and tertiary sites of the alkane. These studies dramatically impact models of alkane oxidation and engine knock (see Box B). Also, the measured differences in reactivity between C-H and C-D bonds provide a rigorous test of present theories of chemical reaction rates.

OH reactions with alkenes (see Figure 2) are complex. We find at T≤500 K, the addition of OH to the alkene double bond dominates reactivity:

$$OH + C_nH_{2n} \rightarrow HOC\ {}_nH_{2n\cdot}$$

The bond holding OH to the alkene is relatively weak, and above 500 K thermal energy is sufficient to begin breaking it. Between 500 and 625 K, the above equation reaction and its reverse approach chemical equilibrium on the millisecond time scale of our experiments with the result that the net rate of chemical reaction decreases as T → 625 K. At still higher temperatures, we measure an increase in the OH + alkene reaction rate when hydrogen-atom abstraction, which has an appreciable rate only at relatively high temperatures, begins to consume OH. Deuterium substitution in the alkene greatly aids our elucidation of the reaction pathways. Our experiments demonstrate that a given reaction may proceed by entirely different pathways in different temperature regimes.

OH reactions with alcohols, $C_nH_{2n+1}OH$, have characteristics similar to both classes of reactions described above. The primary pathway is hydrogen-atom abstraction by OH from the alcohol. Here again, however, the specific site of reaction is important. For OH reacting with ethanol, CH_3CH_2OH, hydrogen-atom abstraction from the CH_3-group produces the species CH_2CH_2OH. This product is the same molecule that is formed when

OH is added to ethene, $CH_2=CH_2$. At elevated temperatures, CH_2CH_2OH dissociates to OH + $CH_2=CH_2$, and the net transformation of these sequential processes is $CH_3CH_2OH \rightarrow H_2O + CH_2=CH_2$. This result demonstrates that alcohol combustion is intimately linked to alkene reactivity. Figure 4 shows the correlation between the reactivities of ethene and ethanol with OH.

Box B
Hydrocarbon Oxidation and Engine Knock

Propane, $H_3CCH_2CH_3$, contains six primary hydrogens and two secondary hydrogens. Primary and secondary C-H bond strengths are different, and OH abstracts these hydrogen atoms at different rates. In propane/air flames, the two possible C_3H_7 radicals that may form in the $OH + C_3H_8 \rightarrow H_2O + C_3H_7$ reaction undergo different subsequent chemistry. OH reaction with the CH_2 group forms H_3CCHCH_3, which then breaks apart to form $H + H_3CCH=CH_2$. Hydrogen atoms are very reactive, and this pathway accelerates the rate of combustion. OH reaction with a CH_3 group produces $H_2CCH_2CH_3$, which leads to formation of the products CH3 + $CH_2=CH_2$. The methyl radical, CH_3, is rather unreactive, and this pathway retards the rate of combustion. Using deuterium substitution methods, we determined the rates of production of the two different C_3H_7 radicals as functions of temperature. These results, displayed in Figure B-1, are critical to accurately computer model propane oxidation. Note that no matter how OH reacts with propane, an alkene is produced when the initial product breaks apart into fragments. Thus it is necessary to understand the oxidation of alkenes and to understand how alkanes burn.

Gasoline for automobile engines is composed principally of hydrocarbons that have chains containing various numbers of primary, secondary, and tertiary C-H sites. Reaction sequences very much like those we have studied for propane lead from the initial fuel molecule to intermediate species that accelerate or inhibit combustion. The fuel octane number relates directly to the distribution among these sites. Recently, Charles Westbrook and William Pitz of Lawrence Livermore National Laboratory have made significant advances in understanding engine knock. They have developed a sophisticated computer model that incorporates our results into a detailed chemical mechanism for combustion. Their work helps quantify how the propensity to knock is connected to the molecular structure of the fuel.

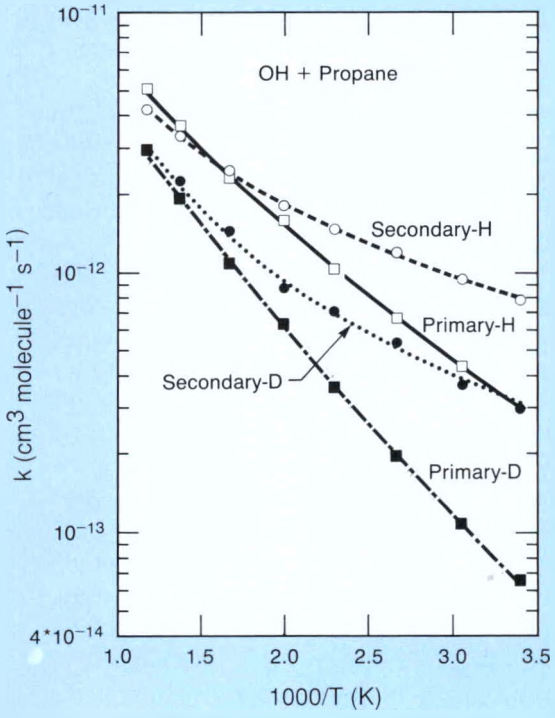


Figure B-1. This figure shows the site-specific rate coefficients for H- and D-atom abstraction by OH from primary and secondary reactive sites in propane. The results shown here allow prediction of the relative amounts of inhibiting and promoting species formed when OH attacks this fuel.

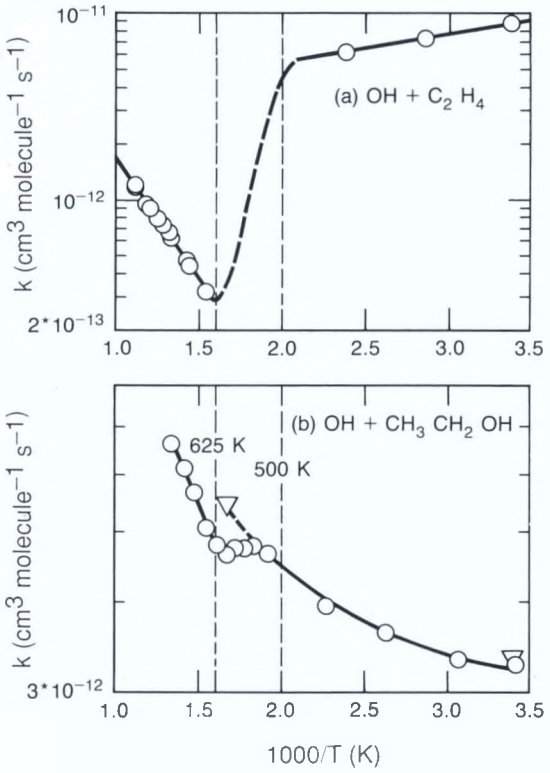


Figure 4
These graphs show the chemical kinetics of the reactions of OH with ethene (a) and ethanol (b). Open symbols show experimental results and lines are empirical fits. Reciprocal temperature is given on the abscissa, and temperature increases toward the left of the graph. For OH + C_2H_4 addition, approach to chemical equilibrium, and hydrogen-atom abstraction reactions define the complex structure of the plot. For OH + ethanol, a plateau in reactivity occurs from 500-625 K, as the common intermediate CH_2CH_2OH decomposes to C_2H_4 + OH. Experiments using isotopically labelled oxygen (triangles) show that the rate of the initial step, hydrogen abstraction, increases monotonically with temperature.

We are using our new knowledge to estimate product yields and to model pre-ignition knock in engines.

We obtain important kinetic and product-yield information using the LP-LIF technique. Our kinetic data on site-specific reactivity in alkanes forms the basis for general formulas that estimate product yields obtained when large hydrocarbon fuels burn. These data also permit realistic modeling of pre-ignition knock in engines. Alkenes form when alkanes or alcohols are burned, and understanding their complex reactions with OH is necessary for combustion modeling. Our rate measurements on the reaction of OH with methanol reconcile previously incompatible low- and high-temperature data, and improve computer simulations of methanol combustion. Our elucidation of the reactive pathways in OH + higher-alcohol reactions, particularly the alcohol → H_2O + alkene conversion route, dramatically alters current models of alcohol combustion.

We expect these chemical studies to become increasingly fruitful in the future. Advances in laser technology will lead to improvements in the kinetic measurements. Also, as computer power increases, we hope a closer and more predictive link between theory and experiment will develop. The information derived from our chemical kinetic experiments serves as an input to generate more accurate computer modeling. Ultimately, computer models containing detailed chemistry will improve the output of and our control over the combustion process.

For more information, please call ***F. P. Tully,*** (415) 294-2316 or ***G. A. Fisk,*** (415) 294-3376.

The Use of Advanced Diagnostics and Computer Modeling to Examine Energetic Material Ignition and Combustion

Applying advanced diagnostics and computer modeling to energetic materials ignition and combustion provides a predictive means to decrease sensitivity and increase safety while maintaining or improving performance.

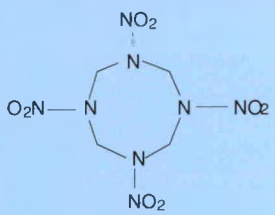
Propellants, explosives, and pyrotechnics, here classified as energetic materials, are an integral part of the huge numbers of munitions maintained within the military inventory. Unfortunately, when energetic material does not behave as expected, serious injury and loss of life can result. Sandia's program to investigate the ignition and combustion of propellants was undertaken when it was recommended that the resources of the DOE National Laboratories be made available for this and other service problems. We are now in the last year of a five-year program with the Department of the Army, and the first year of a program with the Office of Munitions. The aim of our programs is to develop a fundamental understanding of the ignition and combustion of energetic materials. Specifically, we are studying solid and liquid propellants with the expectation that we can provide a more predictive means to decrease sensitivity (thereby increasing safety) while maintaining or improving performance.

To meet this goal, Sandia's Combustion Research Facility uses advanced diagnostics and computer modeling to study/explore energetic material ignition and combustion. We are investigating both solid and liquid propellants and have programs that combine extensive experimental work with state-of-the-art computational modeling. This combination of experimental and computational work provides a unique coupling of research efforts to address very real issues.

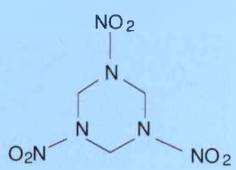
Our program is qualitatively divided into an investigation of two solid propellants and one liquid propellant (see Box A for specific descriptions of each energetic material). Each area of interest is further divided into experimental work and associated computational modeling.

BOX A
Propellant Descriptions

Solids:
Nitramines are a class of compounds characterized by the general formula of R2NNO2, where a nitro group is attached to a nitrogen. The simplest of these is H2NNO2. The energetic material RDX is a more complex nitramine consisting of three nitramine groups bonded together in a ring, while HMX consists of four such groups.



HMX



RDX

Liquids:
The liquid propellant of interest to the Army is a mixture of an oxidizer--hydroxylammonium nitrate (HAN), a fuel– triethanolammonium nitrate (TEAN), and water.

$NH_3\ OHNO_3$

HAN

$(HOCH_2CH_2)_3\ NHNO_3$

TEAN

Why Examine Solid Propellant Combustion?

Solid propellants and explosives have been in use for hundreds of years. Why then are we interested in examining solid-propellant combustion, and what can we offer the military services in terms of improving a technology that has been in existence for so long? Throughout the centuries of energetic material use, people have used trial-and-error methods to generate new compounds, formulas, or catalysts. This rather haphazard approach has worked to some extent in the past, but for present day needs it becomes a costly and time-consuming process. Furthermore, the burning rate models used by the military often employ global rate constants for gas- and condensed-phase combustion of energetic materials to determine the energy release rate. When the reaction mechanisms and rate constants are not known from independent experiments, they become adjustable parameters. As a result, the models lose their physical significance and can be of little value.

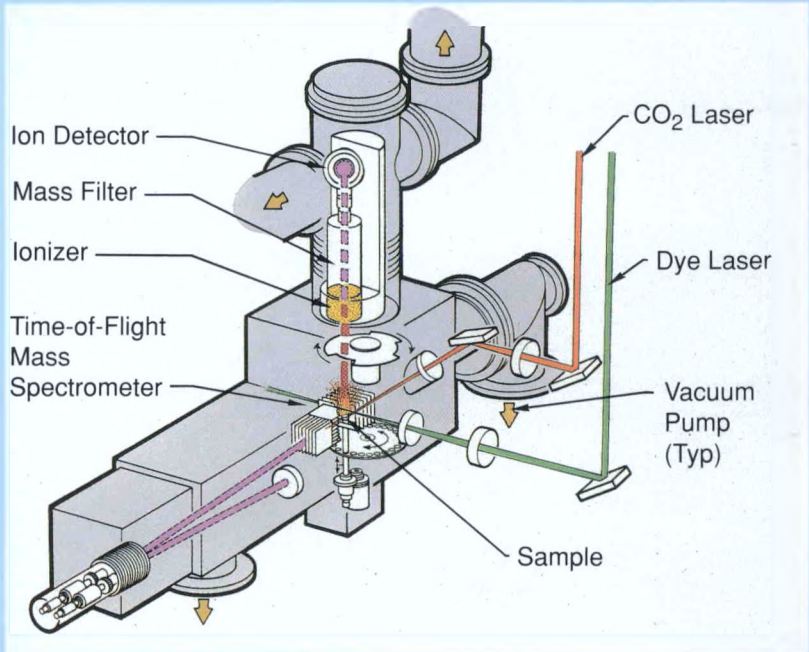
In our work on solid-propellant combustion, we view the burning propellant as a sequence of reaction conditions that are combined to give an overall effect. In this situation, the propellant is made up of a low-temperature, condensed phase reaction zone that generates small fuel and oxidizer molecules as the propellant decomposes. These small molecules enter a primary flame zone where the molecules from thermal decomposition burn to smaller molecules. Finally, the secondary, or luminous, flame zone brings the combustion reactants to their final products and temperatures.

We use a special apparatus to study thermal decomposition in energetic materials.

We built a specialized apparatus (see Box B) to study the decomposition products of energetic materials. Employing this apparatus and heating only small

BOX B
The Study of Decomposition Products

The simultaneous, thermogravimetric modulated beam mass spectrometer shown here examines the thermal decomposition of small samples (10 to 50 mg). The sample is placed in a container and heated to begin decomposition to gas-phase products. Samples can be heated radiatively to provide isothermal or slow heating rate conditions, or a CO_2 laser can be used to obtain very fast heating rates (up to 1000°C/s). Between the time when gas molecules exit the reaction cell and a fraction of them are collected by the mass spectrometer, a reduced back pressure ensures that they do not undergo further collisions. The mass spectrometer consists of the ionizer, mass filter, and ion detector; at the same time, the reduced back pressure is created with the turbomolecular vacuum pumps. The mass spectrometer helps determine the identity and amount of each type of gas decomposition product being formed in the reaction cell as a function of time during the decomposition. Using the beam modulator, we can detect only molecules from the sample container. Also, the instrument allows the approximate molecular weight of the gas products to be determined by time-of-flight velocity spectra measurements.



samples, we are able to determine the identity and rate of release of the gaseous products from thermal decomposition (or pyrolysis) of a solid energetic material. To determine the identity, we use modulated beam mass spectrometry and time-of-flight (TOF) velocity spectra to identify the pyrolysis products, and to measure the rate of gas release from the solid, we use time-dependent measurements of the force change and the mass spectra of the gaseous products as they leave the sample holder under well-established flow conditions. The combination of both the positive identification of the gaseous products from the decomposition along with the determination of their rate of release from the solid leads to a better understanding of the important reaction mechanisms from starting material to final products. These reaction mechanisms are the controlling forces behind the ease of ignition and the rate of combustion. Therefore, if we can alter the reaction mechanisms in a meaningful way, we can then formulate or synthesize tailored propellants.

We examined the thermal decomposition of RDX and HMX at many different temperatures. Only one of our experiments will be discussed here—that is, the isothermal decomposition of HMX at 210°C. The thermal decomposition products from decomposing HMX were measured and identified by careful examination of the mass spectra and TOF velocity spectra at each mass-to-charge (m/z) value. The TOF data are important in assessing the origin of the m/z values obtained from the mass spectra. For example, we have shown that the signals at most of the higher m/z values arise only from fragment ions of HMX and not from its pyrolysis products. Also, the NO_2^+ signal at m/z = 46 arises from the fragment ions of two different pyrolysis products and not NO_2. These results aid us in assessing the reaction pathways from starting material, to intermediates, and finally to products.

Although the mass spectra and TOF velocity spectra measurements are very useful for identifying the pyrolysis products, even these measurements do not always yield unambiguous results. For example, the molecular ion at m/z = 74 could indicate that methylenenitramine, CH_2NNO_2, the monomer unit of the HMX ring, is one of the pyrolysis products. This is a reasonable interpretation based on intuition and the interpretation proposed by previous researchers. However, in our experiments with deuterated and ^{15}N-labelled HMX, we determined the stoichiometry to be $C_2H_6N_2O$, dimethylnitrosamine. This experiment points out how results can be misinterpreted unless proper experiments are conducted.

Unlike previous mass spectrometry measurements on HMX, we have identified the origin of all the major ion signals formed from HMX and the temporal behavior of the pyrolysis products during an isothermal decomposition. Figure 1 shows the rate of gas formation from the decomposition products of HMX and illustrates some of the interesting features.

First, the vapor pressure of HMX rises to a constant value that is maintained throughout the decomposition. This indicates that there is little self-heating of the sample (self-heating would result in a steady increase in the HMX vapor pressure) and that the HMX decomposition products have little effect on the HMX vapor pressure. However, it is possible that both these effects are significant but in opposite directions, cancelling each other out. Thus, the net effect would be zero.

Secondly, the isothermal decomposition results show what has often been described as autocatalytic

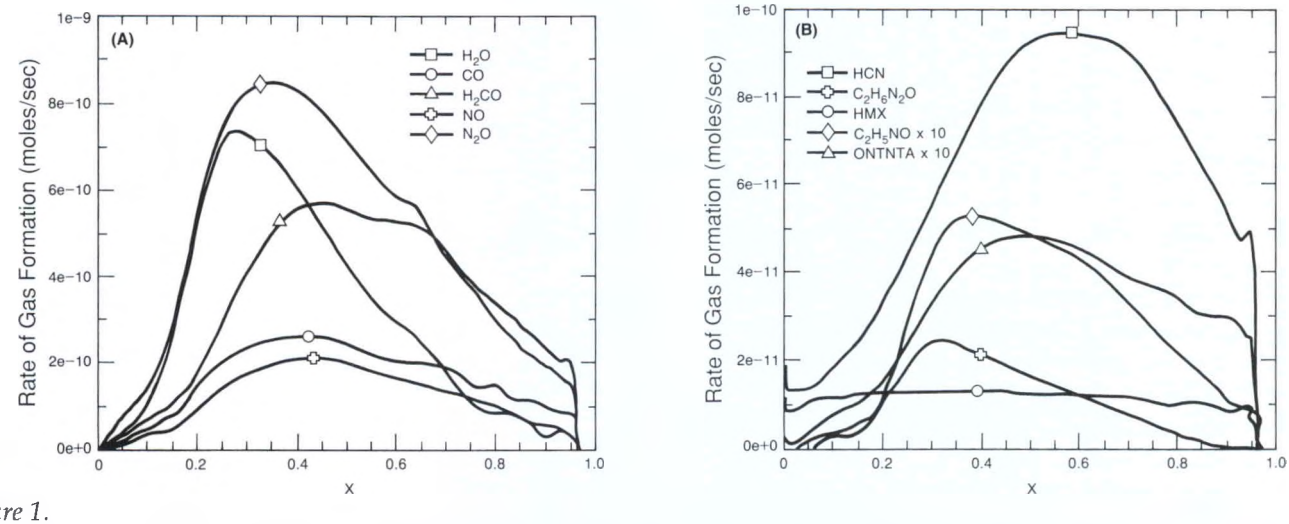


Figure 1.
These figures show the rate of gas formation (moles/sec) versus the fraction (X) of the extent of reaction. By measuring the rate of gas formation, we are able to determine the way that a solid propellant unravels. Once decomposition mechanisms are known, means are possible to alter these mechanisms. ONTNTA is octahydro-1-nitroso-3,5,7-trinitro-1,3,5,7-tetrazocene.

behavior; that is, the results in Figure 1 and in weight loss data show an induction period of relatively slow decomposition, followed by an acceleratory period of more rapid decomposition, and finally a decay stage where the rate of weight loss or gas formation declines. Although autocatalysis may be occurring, our ability to measure time-dependent rates of gas formation opens up the possibility of at least two other explanations: a) The formation of reaction nuclei and structural changes in the solid similar to solid-phase decomposition in inorganic compounds or b) decomposition within the solid propellant may be forming bubbles of gas that release during the acceleratory period. (Supporting this hypothesis, we have micrographic evidence of bubble formation in residues taken from these experiments.) These are all plausible explanations of the behavior of HMX in these experiments and, in fact, some combination of these or others may actually be the case.

Finally, during the induction period the gas products N_2O, CH_2O, and NO form slowly. At the end of the induction period and the start of the acceleratory period, H_2O first appears as a pyrolysis product. The appearance of H_2O coincides with the rapid increase of N_2O, CH_2O, NO, and CO. The formation of $C_2H_6N_2O$ also increases during the acceleratory period, but its rate of increase lags that of H_2O. Later in the decomposition, the ion abundances at the m/z values representing ONTNTA, HCN, and methylformamide increase.

We note with interest that, in contrast to our results, H_2O was not observed as a major product in previous mass spectrometry experiments. This may be due to difficulties encountered previously in the transfer of water from the reaction vessel to the analysis instrument. In our experiments, the products go directly from the heated reaction cell to the mass spectrometer without striking any surfaces or other molecules, so the products cannot be lost or altered before they are detected.

In summary, we have assembled an instrument that couples several important diagnostic techniques into a single instrument not duplicated anywhere else. Furthermore, we have had the use of high-quality propellant samples, including isotopically labelled materials, that have provided new insights into HMX decomposition. These resources have given us the means necessary to measure and unambiguously describe the thermal decomposition products of HMX and other energetic materials and puts us in a unique position to support the Army's needs in investigating the fundamentals of solid-propellant ignition and combustion.

We use quantum chemistry to examine chemical reactions that take place within fast reactions.

Experimentally, we measure the velocity and thickness of reaction fronts and determine the species being produced as a reaction proceeds. Furthermore, one can postulate chemical reactions that may be taking place within the reaction zones to produce the species observed. However, with several very rapid reactions taking place simultaneously, the extent to which these individual reactions contribute remains uncertain. These are the classical circumstances in which one turns to theory for help.

At Sandia we have developed a quantum chemistry approach to examine the various chemical reactions that take place within fast reactions. This approach, called the BAC-MP4 method (defined in Box C), involves a quantum mechanical analysis of the electronic structure of molecules. It yields accurate thermochemical properties such as these: heats of formation, bond energies, and free energies. We then use this method to investigate molecular stabilities, reaction mechanisms, rate constants, and spectroscopic properties.

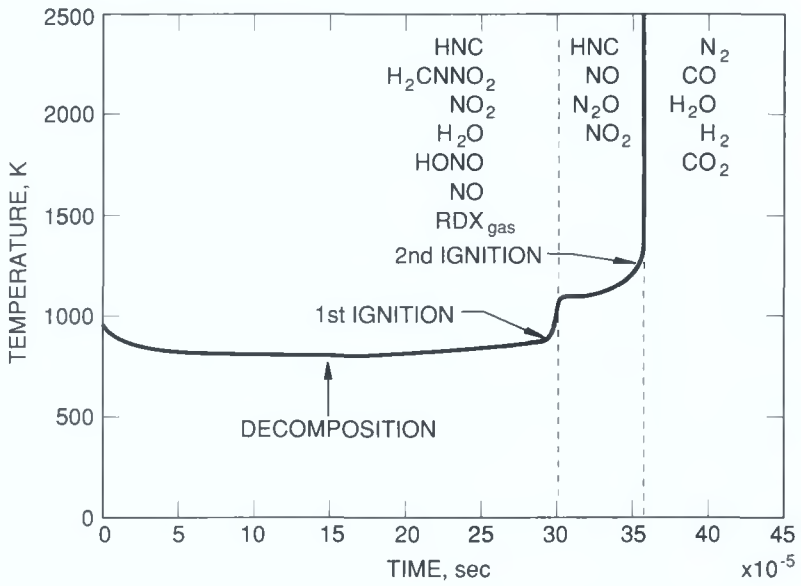


Figure 2.
This graph shows the temperature and chemistry of RDX ignition at 1000 atmospheres. The figure shows an initial decrease in temperature due to the endothermic decomposition of RDX. Two ignition stages are apparent following decomposition. The dominant species during decomposition and each ignition stage are also indicated. From this information relevant reactions can be proposed.

To illustrate the utility of this approach to the investigation of energetic materials, we have examined the ignition process for RDX. We solved time-dependent coupled chemical reactions by using dozens of chemical species and hundreds of chemical reactions. When we examined the results at one atmosphere of constant pressure, we discovered a three-stage process that is consistent with experimental observation. The first endothermic stage (the temperature decreases slightly) represents the decomposition of gaseous RDX. The second stage or first ignition represents the conversion of NO_2 to NO. The third stage or the second ignition corresponds to the major heat release, which forms the final products of combustion. However, one atmosphere of pressure is not representative of real conditions of combustion in solid propellants. To demonstrate how things change at high pressures, we experimentally examined the ignition process of RDX at 1000 atmospheres of pressure (see Figure 2). The resulting temperature profile represents a three-stage ignition identical to the previously mentioned theoretical result. However, the species concentration profiles indicate a change in mechanisms. For example, HONO and H_2CNNO_2 are important intermediates that occur during the first stage ignition but are not predicted as appearing under any circumstances at low pressure.

The BAC-MP4 method is a very powerful theoretical tool to obtain the thermochemical properties of molecular species. The application of the BAC-MP4 method, in providing thermochemical properties of reaction intermediates and identification of reaction pathways, will continue to play an important role in our understanding of propellant combustion chemistry and kinetics.

Determining gas-phase chemistry requires experimental programs to identify formation rate characteristics in propellant combustion.

To evaluate our theoretical modeling effort with regard to gas-phase chemistry, we have implemented two experimental programs to identify reaction intermediates, radicals, and products formed and their rates of formation during propellant combustion. These experimental data are critical in providing our modeling effort with real numbers on which to base predictions. To obtain necessary data, we have two efforts underway: a low-pressure flame experiment and a shock tube/pulsed laser experiment.

The low-pressure flame experiment is designed to examine pairs of reactants that are known to exist in the primary flame zone (or "fizz zone") of a burning propellant. These experiments are performed at low pressure (about 25 torr) so that our experimental reaction zone is spread out spatially to permit identification of the species in the flame. Typically, real propellants burn at much higher pressures and give a reaction zone that is much too thin to diagnose effectively. The experiments begin by burning premixed fuel and oxidizer over a

BOX C
The BAC-MP4 Method

The BAC-MP4 approach begins with an electronic structure calculation of a given molecule using the Hartree-Fock method. This technique is used to determine the optimum molecular geometry and harmonic vibrational frequencies of the molecule. Total electronic energies are then calculated at a higher level of theory using the Møller-Plesset many-body perturbation theory to fourth order (MP4). This method extends Hartree-Fock theory to include electron correlation, important in evaluating bond energies. We then include bond-additive corrections (BAC) to obtain heats of formation. Combining this information with the moments of inertia of the molecule and the vibrational frequencies provides the thermochemical entropies and free energies of the various decomposition and combustion intermediates. For transition state structures, we use the same approach except that we locate a saddle point on the electronic potential energy surface. The table compares the BAC-MP4 and experimental heats of formation for selected molecular species of relevance to energetic materials. We find that the BAC-MP4 method works well for unstable and radical species as well as for the stable molecules.

	$\Delta H^o_{f,298}$			$\Delta H^o_{f,298}$	
Molecule	Theor.	Exp.	Molecule	Theor.	Exp.
CH_3NO_2	-69	-75	HNCO	-116	-112
CH_3ONO	-65	-65	HN_3	301	300
CH_3ONO_2	-110	-120	N_2O	82	82
CH_3NHNH_2	92	85	NO	87	90
NH_2CHO	-192	-186	NO_2	27	33
HONO	-81	-79	CH_3	146	146
$HONO_2$	-128	-135	CH_2	388	386
CH_3CHO	-162	-166	NH_2	192	185
CO	-123	-111	NH	363	343
CO_2	-392	-394	OH	39	39
HC(O)OH	-384	-379			
CH_3OOH	-125	-131			

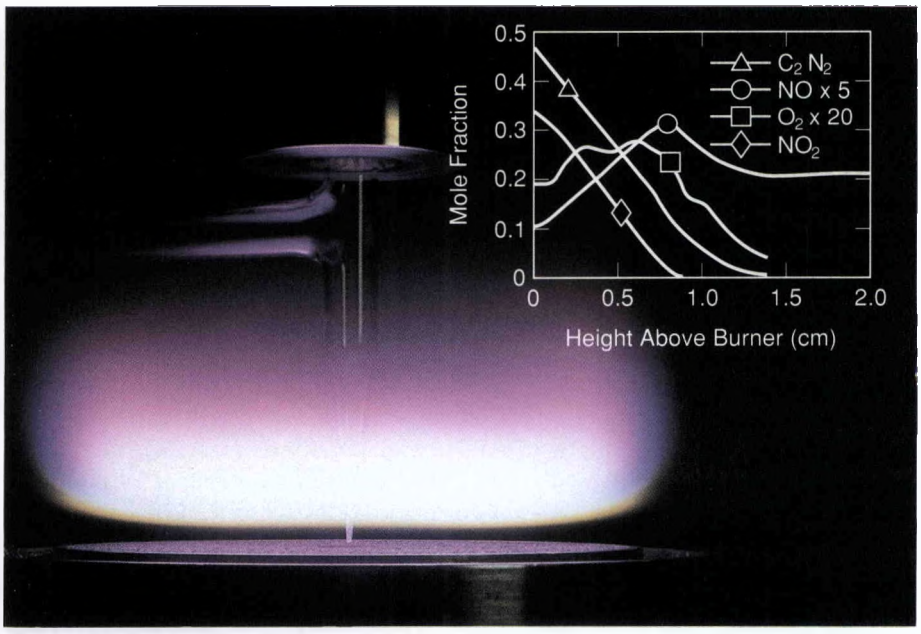


Figure 3.
A cyanogen/nitrogen dioxide flame ($C_2 N_2$ /NO_2) is pictured here. A quartz probe extracts gases at various heights above the burner surface. A mass spectrometer identifies the gaseous species and a typical plot of mole fraction vs. height above burner is shown in the inset.

flat flame burner that can translate up and down. We employ two types of equipment to identify species above the burner surface: a mass spectrometer and a Fourier transform infrared spectrometer. They are complementary techniques and provide quantitative data on species concentrations as a function of the distance above the burner. Figure 3 shows a typical flame with an example of the data collected from the mass spectrometer.

We have made some revealing comparisons to theory with data from experiments on low-pressure flames, but conditions in those experiments are far from actual propellant use. To reach higher pressures and temperatures, we have assembled a shock tube to investigate propellant combustion at more severe conditions. Although shock tubes were invented nearly one hundred years ago, only recently have the necessary diagnostics been available to provide the very fast and highly sensitive detectors required to examine

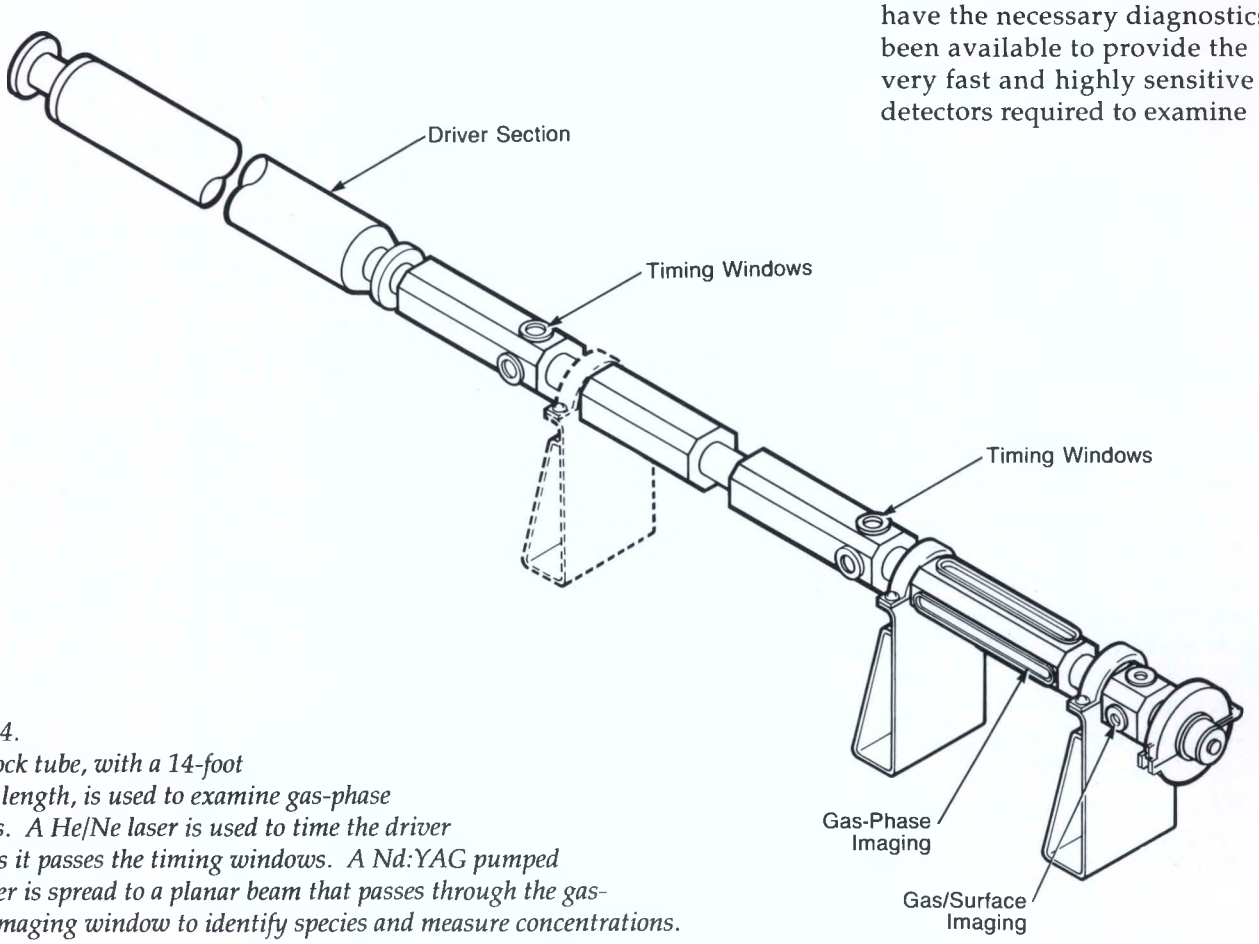


Figure 4.
The shock tube, with a 14-foot overall length, is used to examine gas-phase kinetics. A He/Ne laser is used to time the driver pulse as it passes the timing windows. A Nd:YAG pumped dye laser is spread to a planar beam that passes through the gas-phase imaging window to identify species and measure concentrations.

relevant propellant gas-phase reactions. Figure 4 illustrates the shock tube and the associated methods of detection for chemical species. A typical shock tube experiment examines species produced from premixed binary gas mixtures (e.g., HCN/NO_2). However, we also have a unique capability to deposit a thin film of solid material (e.g., RDX or HMX) on the end wall of the tube, shock the deposited material, and then identify species that are evolved in real time. Furthermore, by codepositing additives with the propellant, we hope to be able to modify and control the reactivity of the propellant and provide a means of maintaining performance and decreasing sensitivity.

We examine solid-propellant combustion by modeling the transient behavior of propellants.

The motivation to experimentally examine shock heated films of RDX and HMX (see preceding section) comes as a result of modeling the transient behavior of propellants described in this section. The Denison and Baum (D&B) model, a classical model commonly used in propellant ignition dynamics, involves a single-step, endothermic solid-to-gas pyrolysis, followed by a single-step, exothermic, gas-phase flame. Both of these reactions must have a high-activation energy for the solid to be stable at room temperature and for the customary thin gas-flame to form. Partly because of this high-activation-energy construction, the steady burning solution is easily calculated and gives a compact, realistic form for the steady burning rate. Unfortunately, this high-activation-energy assumption also causes the time-dependent behavior of the model to be exceedingly stiff (i.e., unresponsive). Such a propellant model is difficult to ignite, and once ignited, is difficult to maintain. Figure 5(A) illustrates this symptom for radiant ignition, an energy source for such a propellant model. At time t = 0, the propellant has been ignited by radiation impinging on the solid surface and continues long enough to establish a steady burning condition. Once we establish a steady burning condition, the radiation is suddenly terminated. The result of using the D&B model in this example is that the propellant extinguishes and never reignites. (The phenomenon in question is not new; it is usually termed deradiative extinction. It serves as a useful benchmark tool for investigating the dynamics of energetic materials.) In the detailed analysis of gas flames, an essential component is low-molecular-weight, low-activation-energy radical reactions. Shown in Figure 5(B) is the same transient as Figure 5(A) for a proposed reaction scheme that involves radical reactions, both at the surface and in the gas phase. The previous high-activation-energy reactions are still present; however, each reaction produces a radical (denoted "A" at the surface and "H" in the gas). The gas-phase radical can diffuse against the flow to react with the surface if it is not first absorbed via recombination (H+A). As in strictly gas-phase

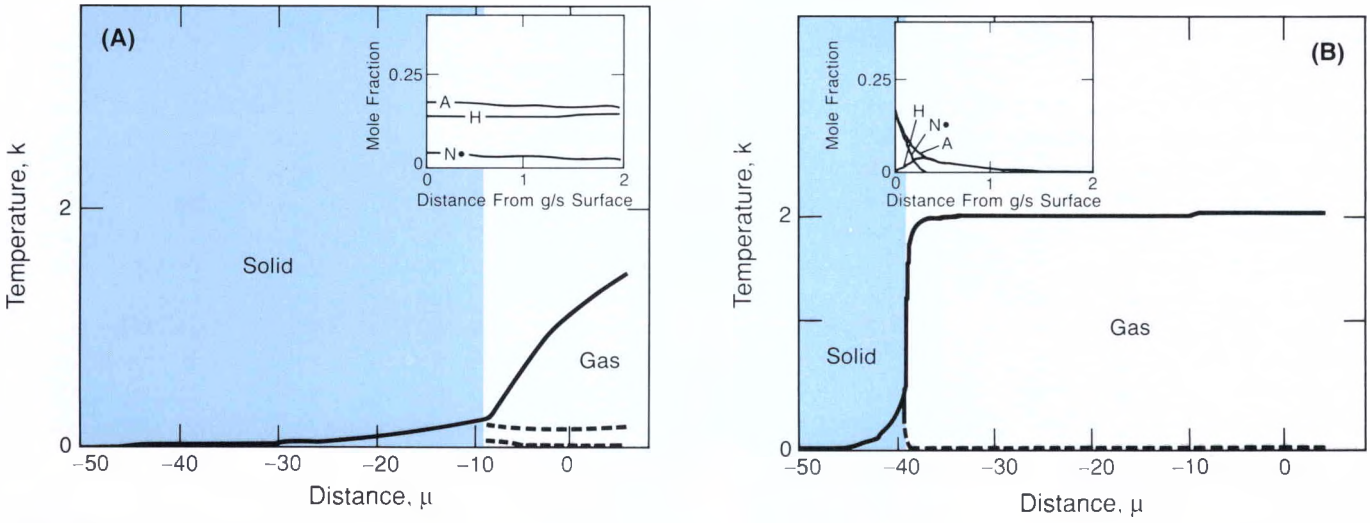


Figure 5.
The ignition characteristics of a solid propellant predicted by our combustion model are shown above. Our model in Figure 5(A) does not sustain combustion after ignition; the solid surface stops moving to the left and the gas temperature falls. After ignition of the model in Figure 5(B), combustion is sustained. Sustained combustion is a result of including molecular fragments of radicals that participate in reactions at the surface of the propellant or in the gas phase. This is seen by the displacement of the solid surface to the left, compared to the model in Figure 5(A), and the temperature remaining high.

combustion, the recombination reaction plays a major role and, in this particular case, ensures that the chemical routes during steady combustion are dominated by the D&B reactions. Thus, the steady burning rate is little changed from that of the original D&B prediction. However, Figure 5(B) shows that the unsteady behavior of this system, containing an insignificant quantity of radicals for the original D&B reactions, makes the difference between a propellant that extinguishes and one that attains a steady burn long after the radiation on the surface has been terminated. Although unconfirmed experimentally, our model indicates that the inherent gain instability due to radicals is tied less to the exothermicity in their surface or recombination reactions than to the liberation of species from the surface that can diffuse to regions of the gas still hot enough to allow them to react.

In any case, it is clear that models, such as the D&B model, that do a good job of predicting steady behavior in energetic materials may be incapable of predicting the outcome of simple transients.

All propellant models have at least two solutions: one burning, one not. We have shown that a change in chemistry (i.e., the inclusion of radicals) does not affect the burning rate of the propellant but does dramatically affect propellant response to an ignition source. In other words, the propellant ignites and continues burning even after the ignition source is removed–the real behavior of propellants.

Can We Replace Solid Propellants With Liquid Propellants?

The U.S. Army is seriously considering liquid propellants to replace existing solid propellants in certain artillery and tank gun applications. The liquid monopropellant that is of interest to the Army is a mixture of an oxidizer—hydroxylammonium nitrate (HAN), a fuel—triethanolammonium nitrate (TEAN), and water. This propellant is an insensitive energetic material needing very high pressures (about 1000 psi) before it will sustain combustion. Thus, the propellant could be stored at atmospheric pressure, and there would be no danger to personnel from an unexpected explosion or detonation. Furthermore, compared to a solid propellant, the liquid propellant is roughly a tenth the cost, has a higher volumetric energy density, and provides more rapid firing, resulting in more munitions on target. However, liquid propellant has some disadvantages that we have considered in our research. For example, liquid propellant cannot be fired in a bulk-loaded configuration like solid propellants; the burning of the liquid is too unstable, resulting in explosive destruction of a standard gun. To prevent such destruction, a regenerative 155-mm howitzer gun configuration has been designed and is now being successfully tested (see Figure 6). Although muzzle velocities are comparable to solid propellants, pressure oscillations within the combustion chamber are large, often half again that of the mean pressure. These pressure oscillations seriously concern the Army because of the erosion of hardware and the resultant reduced component lifetime as well as the unknown behavior of the many Army projectiles to unexpected and unreproducible oscillations. Our work in liquid propellant combustion is focused on understanding the oscillations, the burning rates of propellants, and the means to eliminate or reduce the oscillations. Our goal is to make liquid propellants a viable candidate for the next generation propellant in artillery and tank guns for the Army.

Understanding the role of combustion instabilities in liquid propellants is the goal of our studies.

If we are to address the concerns of pressure oscillations observed in prototype artillery hardware, it is imperative that we understand the possible causes of instability. Specifically, the role of combustion instabilities is of primary concern. For a liquid propellant, these instabilities are a result of both the effect of pressure on the planar burning rate and to instabilities inherent in the combustion of liquids. To investigate these issues, we have established a facility that is uniquely suited to the study of liquid propellant combustion at ignition pressures (up to 5000 psi). The facility consists of a high-pressure-reaction vessel in which we can study the combustion of up to 100 milliliters of propellant. We employ various diagnostics to study the chemical and physical processes relevant to propellant combustion.

The monopropellants being considered for use in liquid propellant guns are different from other liquid propellants in several respects. The liquids are highly ionic solutions that contain approximately 20 percent water. Thus the chemistry of liquid propellant combustion may be affected by ionic chemistry and, at the pressures of interest, by the presence of supercritical water. One of the most

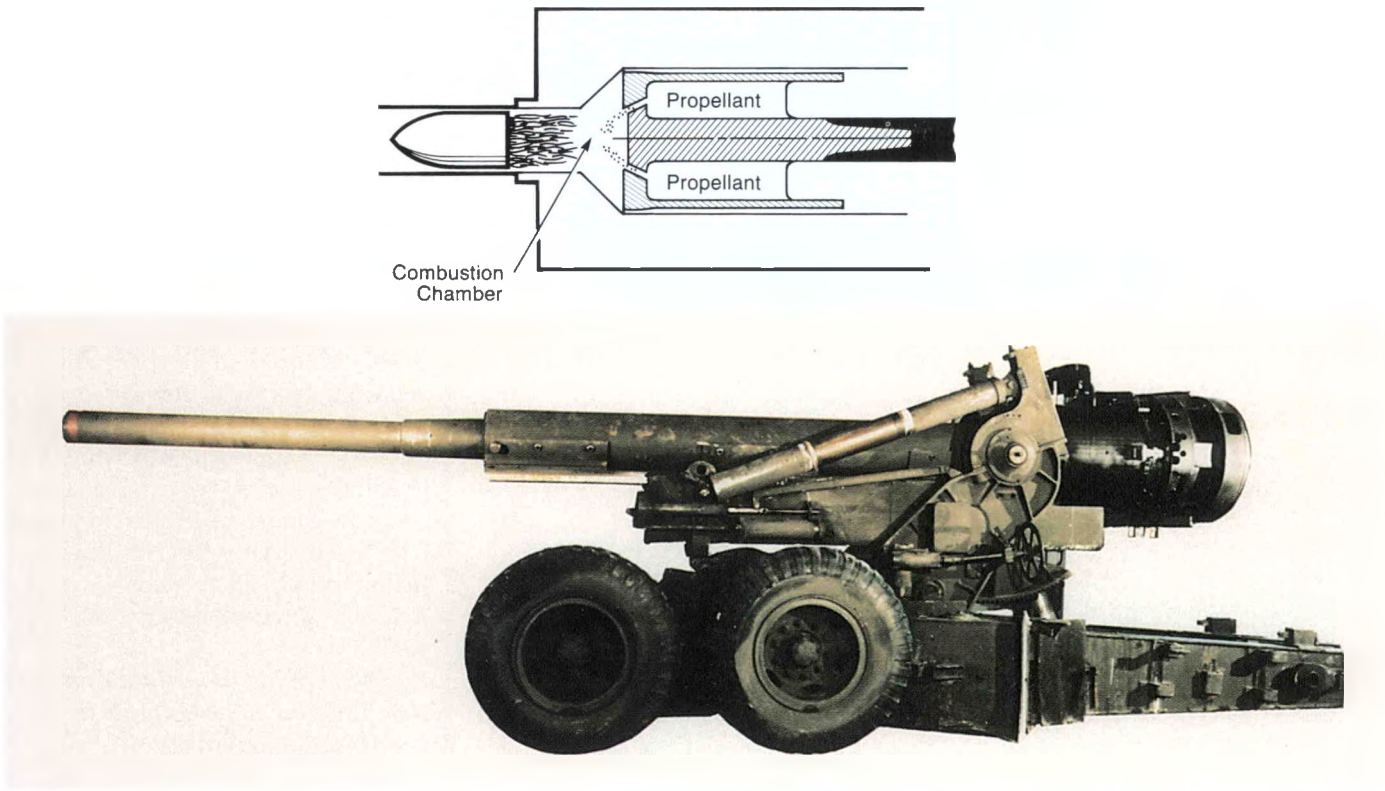


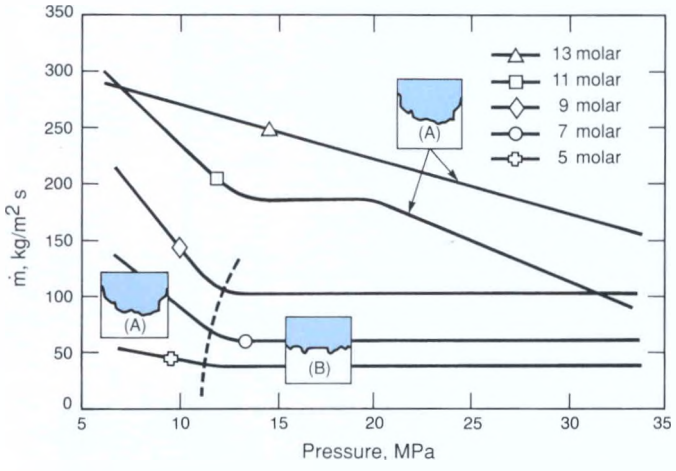
Figure 6.
This prototype 155-mm regenerative liquid propellant gun (RLPG) is being tested by General Electric. A regenerative design is necessary when using LP1846, a ratio of HAN to TEAN that is about 20 weight percent water. This mixture maintains control of the combustion process and prevents unwanted destruction of the breech and combustion chamber. The inset illustrates the workings of the RLPG.

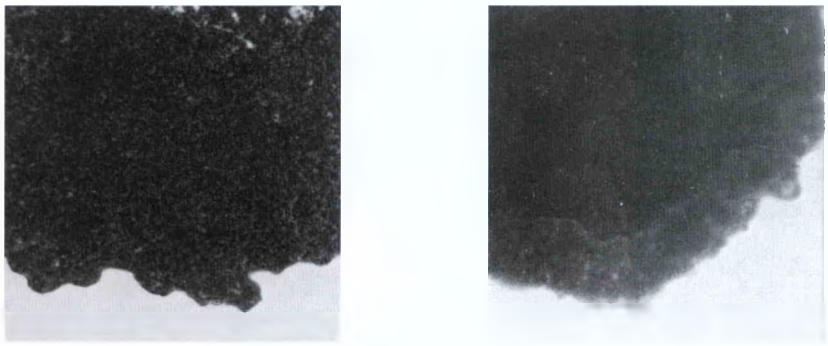
interesting properties of these propellants is that, unlike other liquid propellants, the combustion is dominated by condensed phase reactions (i.e., the thermal decomposition of HAN). The importance of condensed phase reactions on the burning rate and stability of the liquid decomposition zone are current areas of research.

Initial work on the combustion of liquid propellant confirmed the importance of HAN decomposition on liquid propellant combustion and that the overall combustion rate decreases with an increase of pressure. In addition, we found that the HAN decomposition rate was the rate limiting step in propellant combustion and that HAN-water mixtures decompose at speeds similar to the burning speed of the full propellant formulation.

To investigate the effect of HAN on propellant decomposition, we conducted a series of experiments to determine the decomposition rate of HAN-water mixtures over a range of pressures (1000 to 5000 psi) and HAN concentrations (3 to 13 molar). Figure 7 shows the apparent decomposition rate as a function of pressure for several HAN concentrations. For each mixture, the decomposition rate is seen to decrease with increased pressure, as was noted for liquid propellant combustion. We found that the decomposition interface was composed of instabilities of two wavelengths. These small scale instabilities—about 50 micrometers across—gave the interface a corrugated appearance that did not change appreciably with pressure. The effect of pressure was most noticeable on the larger wavelengths (on the order of several millimeters) that decreased in amplitude with an increase of pressure. These instabilities have been attributed to hydrodynamic instabilities across the liquid gas interface. We found that the change in surface area due to the larger waves is apparently responsible for the variation of decomposition rate with pressure.

Hydrodynamic instabilities have a great effect on the decomposition rate of HAN-water mixtures, and thus on the burning rate of liquid propellant. These instabilities—on the order of several millimeters—are most certainly of importance in liquid propellant combustion in gun environments over the pressure





(A) (B)

Figure 7.
This figure shows the apparent decomposition rate M of HAN/water mixtures as a function of pressure. Two wavelengths of instabilities, detailed in the boxes on the graph, are shown in full-size photographs (taken at 5000 frames/sec). The first are small corrugations on the liquid surface, seen in both photographs. The second are larger and are seen as a meniscus in photograph (A) but disappear as the pressure increases, seen in photograph (B).

regime investigated. This work will be instrumental in guiding gun designs to optimize performance of this new insensitive propellant.

Our liquid propellant injector/combustor device provides state-of-the-art diagnostics of combustion behavior.

The strand burning experiments described above have proved valuable in obtaining a fundamental understanding of the chemical and physical processes pertinent to the early stages of liquid propellant ignition and combustion when pressures are still relatively low. However, the pressure regime in those experiments is limited to a tenth of that developed at the peak of combustion in a real system; thus, average pressures of 50,000 psi are observed in prototype hardware, and this value established the need for an experimental program to apply sophisticated diagnostics to a gun-like device. To meet this goal, we have designed and built a liquid propellant injector/combustor (Figure 8).

Our device uniquely provides state-of-the-art diagnostics to learn the origin of the oscillatory pressure behavior during combustion, a means to eliminate or reduce the oscillations, and a tool to observe the liquid breakup. Currently, no other researchers are applying a systematic approach to understanding the pressure oscillations; they are using trial-and-error methods with diminishing success. For propellant breakup, gun designers and modelers assume that the liquid is distributed into 100 μm droplets, and then they calculate a certain burning profile. However, the burning profile is not that observed in prototype hardware. By learning the origins of the oscillations and providing a description of propellant breakup, we expect to provide the information that is necessary to optimize gun performance.

Modeling liquid propellant combustion plays an important role in our experimental program.

Modeling of liquid propellant combustion plays an important role in understanding the significance of our experimental programs. We have two efforts that model liquid propellant behavior: (1) the development of a dynamic model to investigate the origin of the instabilities observed in the strand burner experiments and (2) a decomposition/reaction model for LP combustion.

Our dynamic model of the strand burner experiments centers on the forces which discourage or enhance steady, planar burning at the liquid/gas interface. Previous investigations have not coupled reactive/thermal-diffusive interactions (thermochemistry) with inertial/viscous interactions (fluid mechanics). Our goal is to merge these two pictures, produce a comprehensive stability diagram, and learn how liquid disturbances are created. We have now produced a stability diagram corresponding to typical liquid propellant burning conditions and identified regions of reactive/thermal-diffusive and inertial/viscous stability. At this time, we are examining the experimental results

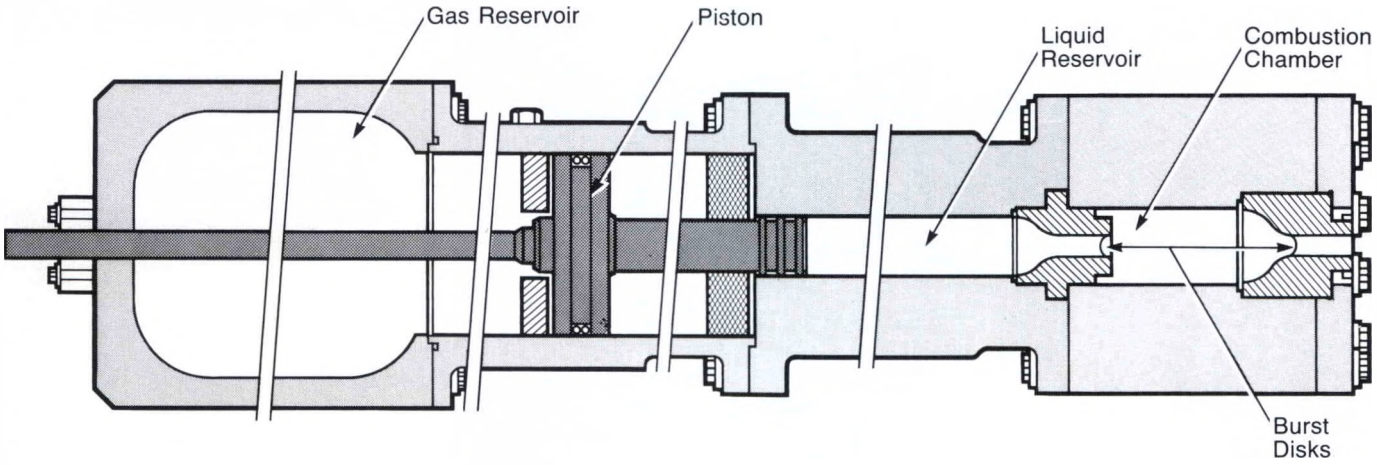


Figure 8.
A cross section of our liquid propellant injector/combustor (LPI/C) indicates the important components. The LPI/C is designed to combust LP1846 at 50,000 psi in less than 10 msec. State-of-the-art diagnostics (not shown) are employed to examine liquid breakup and combustion to determine the origin of unwanted pressure oscillations.

further to understand how stable and unstable regions impact the burning rate, especially at lower pressures (ca. 1000 psi). Also, we are conducting experiments to better resolve individual disturbances to understand the origin of these disturbances (that is, are the disturbances caused by reactive/thermal diffusive, inertial/viscous forces, or other yet undetermined factors).

The decomposition/reaction model is intended to produce a general tool to study liquid-propellant combustion parametrically. A model that covers a wide range of pressures and chemical compositions is required to understand experiments that have been carried out in strand burners and in droplet form. The modeling results have shown that HAN decomposition is the rate-determining step (consistent with experimental observations) and that the reaction zone is so thin that finding an experimental means to investigate the reaction zone at realistic pressures will be very difficult.

Both these modeling efforts are having an impact on the directions we are taking in our experimental programs through which a clearer picture of liquid propellant ignition and combustion will be developed. From these collected efforts, we expect to show that liquid propellants are viable alternatives to solid propellants and may be used in future applications.

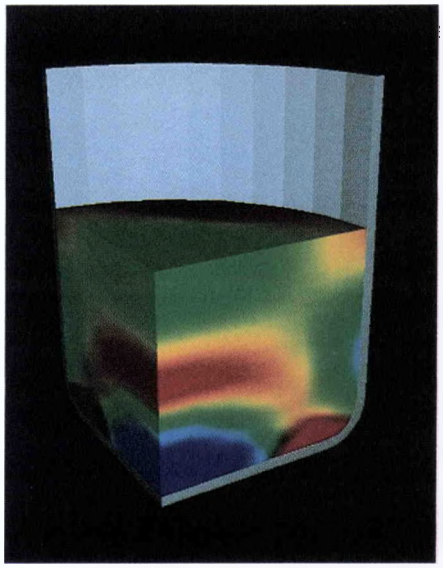
This research has resulted in significant improvement in our understanding of the combustion of energetic materials.

Our long-term objective for these propellant programs is to improve our understanding of propellant combustion, and to this end we have made important observations. We have resolved the rate of gas formation from the thermal decomposition of HMX. Our flame code has successfully identified a two-stage flame at high pressures for RDX and identified the important gas-phase species. We have also developed a thermochemical model that successfully characterizes the transient behavior of propellants. Our modeling has played an integral role in directing additional experiments to examine gas-phase reactions in a low-pressure flame and a shock tube. This work uniquely supports the Army's investigation of the fundamentals of solid-propellant ignition and combustion.

In the liquid propellant portion of the program, we have identified HAN decomposition as the controlling ingredient in the combustion of the propellant and have developed a model that will distinguish the origin of the instabilities observed in experiments.

Sandia has an unusual combination of state-of-the-art experimental and computational capabilities that can be applied to the ignition and combustion of solid and liquid propellants. Our contributions will be the basis of our future work and should play a significant role in optimizing safe propellant performance, and providing guidance for the synthesis and formulation of new energetic materials.

For more information, please call ***Bob Carling***, (415) 294-2206.



This computer-generated color image shows a three-dimensional finite element simulation of a forging process. The quarter section shows the internal distribution of deformations after the billet has been pushed downward into the cylindrical die. The red regions correspond to the highest deformation level and also to the white regions in the micrograph of a sectioned forging shown in Figure 5.

Computer Simulation of Metal Forming

Computer-aided modeling and analysis of deformation processing offers a cost-effective means of enhancing, as well as creating, fabrication technologies for advanced engineered materials.

Since the time of early civilized man, metal shaping and forming have been instrumental parts of the technology of many societies. Early Western and Eastern cultures shaped metals to make tools for agriculture and weapons for battle. In competing medieval European societies that lived by the sword, the superiority of one faction over another could be established by better metal forming technology with which to manufacture swords and armor of superior strength. The importance of metal forming technology continues into this century: we still rely upon metals for many of our technologies, including, just as in centuries past, the manufacture of weapons. However, in today's world, the pressures for superior metal forming technology are driven by more complex global, economic, and political factors.

Today's challenges in advancing metal forming technology are similar to those of the past but revolve around customizing shapes and properties of more advanced alloys, using more sophisticated techniques. Fortunately, we have available a tool for advancing our technology that was not available to our ancestors—the computer. Sandia scientists and engineers have taken a leadership role in advancing metal forming technology by showing how we can use computer simulations of metal forming processes to predict the properties of metal parts without ever manufacturing anything.

Traditional methods of metal forming have become less and less acceptable.

Metal forming encompasses a broad range of manufacturing methods that accomplish in different ways at least one or both of two principal goals: to obtain metal products with particular shapes and to alter the intrinsic properties of metals by the combined effects of deformation and heat. Some of the better known methods for obtaining these goals include rolling thick billets into thinner sheets, extruding rods into wire, and forging billets into dies with custom shapes. (The term billet refers to a block, plate, or rod of metal in a rough condition compared to its final condition.) The latter example, forging, is especially challenging because both goals must be met to obtain special shapes by forcing metal into dies and to impart special properties to the shaped metal. Ironically, the most common means for obtaining these goals is to try out a particular method and see if it works. If it does not, then the method is modified and tried again and again until it is successful.

In many cases this trial-and-error method for developing metal forming processes is adequate. However, this method is becoming less acceptable as we increasingly demand better performance of traditional alloys and use more advanced engineered alloys. The traditional trial-and-error approach may limit product quality, increase cost, and even prohibit the use of advanced materials. Our concern about cost and quality is amplified in the United States by some examples in our metals industry that do not meet the production standards of our foreign competitors. Although Sandia has no direct responsibility for products that are offered in the world markets, we are responsible for designing and developing weapons for national security and for providing technology to address other national priorities.

Computer technology is leading the next manufacturing revolution.

For years, Sandia scientists have relied upon the capabilities of computers to aid in engineering design and scientific research. Now we are demonstrating how computers can aid the manufacture of forgings. The potential benefits of using computer models to simulate forging processes are substantial—enough to revolutionize design and manufacturing methods.

Before we describe our computer simulations, it would be helpful to understand the fundamentals of metal forming process development and, in particular, forging. Typically, a forging procedure is "developed" when manufacturing engineers receive a request to make a newly designed metal part. Figure 1 shows an example of such a forging. In the Department of Energy's Nuclear Weapons Complex, new designs are generated by engineers at Sandia, and also at other national laboratories, and then distributed to one of several DOE-operated manufacturing facilities or to commercial metal forming facilities. The engineers at the manufacturing facilities are responsible for developing a process that will meet the requirements specified by the part designer. This specification will include a range of characteristics, including the type of metal required, the final shape, and the metal's properties, most of which are determined by the microstructure of the metal. Microstructure refers to the complex details of a metal's characteristics as they appear on the scale of dimensions ranging from the diameter of a human hair down to the size of a cluster of a few atoms. By carefully selecting the sequences of deformation and heating during a forging process, we can customize microstructures to obtain the required materials properties. Our purpose then is to use computers to simulate forging processes and to determine what sequences of shaping and heating result in the proper final shape and in the properties that best meet the requirements specified by the design engineer.

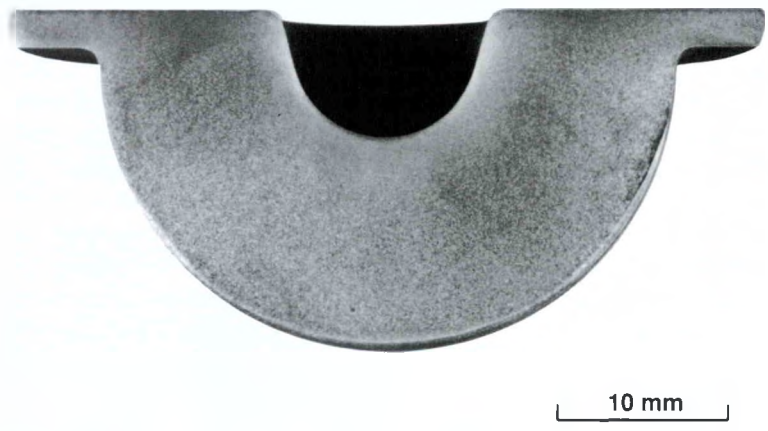


Figure 1.
This micrograph shows a cross section of a thick-wall hemispherical forging produced by a two-stage process.

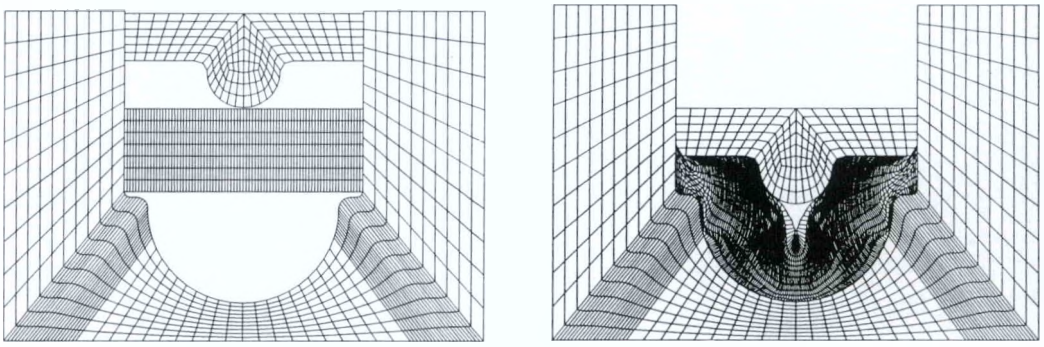


Figure 2. Finite element meshes at a beginning and an intermediate state show that the pictured single-stage process intended to obtain the hemispherical shape shown in Figure 1 from a rectangular block of metal cannot be achieved without forming the undesirable gap that appears under the punch in (B).

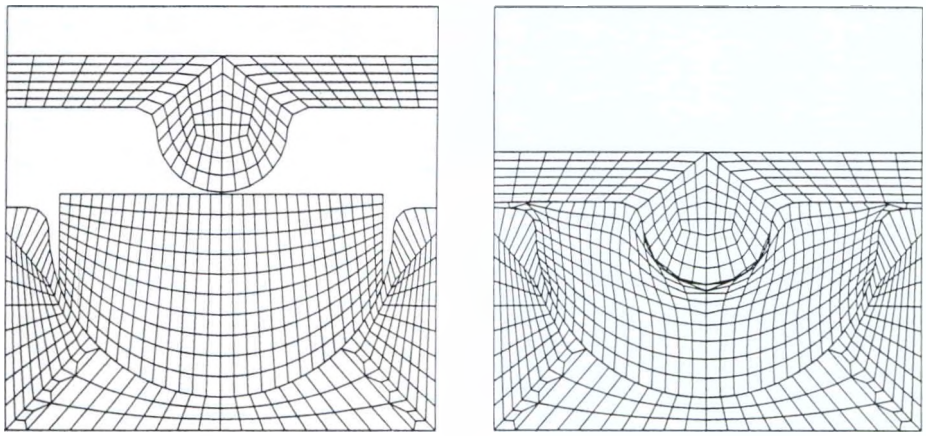
The challenge of using computational models to assist in forging production is multifaceted. To develop the needed technology successfully, we must integrate several scientific and engineering specialties, including mathematics, continuum mechanics, metallurgy, and mechanical testing. To implement computer simulation into forging process development, we must coordinate the efforts of designers, skilled analysts, manufacturing engineers, and quality assurance personnel. Our work has been aimed at providing several pieces of the required technology. These pieces revolve around the application of a computational method known as the Finite-Element Method or Finite-Element Analysis. This method has been used for many years to analyze the deformation of structures (e.g., bridges, airplane wings) but had not been used extensively to simulate metal forming, except in isolated applications to simple problems. (The meshes in Figure 2 show the gap that develops during this process.) Our objective has been to determine whether existing finite-element computer codes can help us produce high-quality forgings, on schedule, and at minimum cost.

Finite-element modeling of production forging processes is fast and economical.

To begin our work, we selected two finite-element codes, NIKE and DYNA, that contain features suitable for analysis of metal forming. Using these codes was particularly simple because of the availability of the simple but powerful companion programs—MAZE to prepare their inputs and ORION or TAURUS to graphically visualize their output. All of these programs were developed at Lawrence Livermore National Laboratories and run efficiently on the CRAY computers that were used for our simulations.

We selected several existing production forging process development problems to analyze. It was imperative that any analysis be completed within the period of several months that is typically permitted for process development and initial production trials. Because of the ease of using MAZE, ORION, and TAURUS, the time-intensive steps of generating the input to the finite-element codes and analyzing the calculated results seldom required more than several days for each analysis. In addition, no single analysis, once properly set up, required more than 15 minutes of CRAY-1S supercomputer central processor time to complete. Most two-dimensional calculations required only two to three minutes. Using a faster and larger CRAY X-MP24 supercomputer reduced these times by 30%. Multiple calculations using different process parameters could be performed quickly and economically. Computation times in trial calculations performed using a minicomputer, a VAX 8700, were approximately nine times longer than for the CRAY-1S. However, we were able to clearly show that computer performance has reached a point where the major cost to obtain useful results is the labor of skilled analysts, not the cost of computing. Furthermore, typical problems could be analyzed within days—a major time savings over trial-and-error production methods.

But what were the benefits? These were seen by examining two typical problems—a two-dimensional analysis of a hemispherical forging and a three-dimensional analysis of a cylindrical forging.



(A) (B)

Figure 3
These views of meshes show the beginning (A) and final (B) states of the second stage of a two-stage process to obtain a hemispherical forging. The preshaped billet that is shown resting in the die in (A) was obtained by a first-stage process. The final shape obtained in (B) had precisely the required shape.

Case histories show the benefits of using computers to simulate forging processes.

Hemispherical Forging

Our first analysis was performed to help design the process for producing the thick-wall hemispherical forging shown in cross section in Figure 1. We addressed two issues: 1) whether a single-stage or two-stage forging sequence would be necessary and 2) what processing conditions would be necessary to obtain specified mechanical properties. The first objective, to choose between a single-stage or two-stage process, was largely a matter of determining the patterns of metal flow during forging. This objective was readily achieved by simulating metal flow for several alternative die geometries. Within a week we showed that it was not possible to forge a rectangular billet into the desired shape in a single stage. The distorted mesh in Figure 2 shows the gap that inevitably forms beneath the center of the punch as the metal is driven downward into the die. We then turned to a two-stage process, the second stage of which is shown in Figure 3. In the first stage (not shown), the rectangular billet was forged to have the preform geometry shown in Figure 3(A). Using the same die as in the first stage, the preform was then pressed into the final shape shown in Figure 3(B). This operation was successful and was used in full-scale production of this forging. However, the analyses pointed out that the second objective, to obtain certain mechanical properties, would be troublesome.

Our anticipation of trouble materialized after we plotted the distribution of deformation as shown in Figure 4. This contour plot shows that very little deformation occurs near the outer radius of the forging. From prior experience in forging this particular alloy, we knew that only by applying large enough deformations could we modify the microstructure of the metal enough to obtain the required minimum strength. Subsequent tests confirmed that, indeed, the strength of the metal in regions near the outer radius was too low. However, by relying on the knowledge of the strain gradients predicted by the finite-element analysis, we obtained a compromise solution by machining the final part from regions of metal that best met the designer's requirements.

The direct benefits of this simulation demonstrated in advance of manufacturing trials a likely forging sequence to obtain the desired final shape and to anticipate potential difficulty in achieving the required strength. However, of greater general significance was the successful demonstration that computer simulation can aid manufacturing engineers in producing forgings with less development time and more confidence in the quality of the final product.

Cylindrical Forging

The cylindrical forging analysis was performed to address a specific problem that occurred during the trial-and-error development of a process to high-energy-rate-forge a rectangular billet of stainless steel plate into a short cylindrical preform. The preform was intended to have a uniform microstructure, but micrographs of sectioned forgings from a trial production lot showed substantial inhomogeneity, as seen in Figure 5.

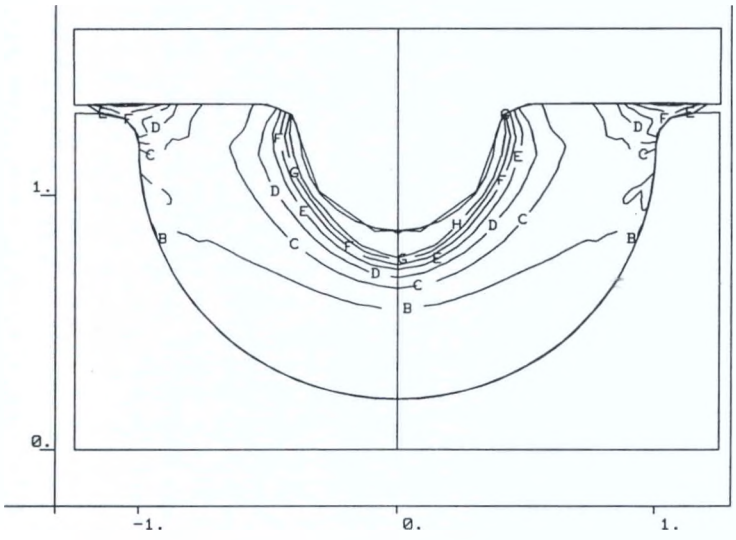


Figure 4.
The levels of deformation (represented here as contours of effective plastic strain) show that much more deformation occurs near the punch than at the outer radius of the forging. The strains in the region between the B contour and the radius were smaller than desired.

We hoped that by simulating the process used in the production trials, an alternative process could be proposed to obtain a more uniform microstructure. We set up two-dimensional and three-dimensional analyses and computed the distributions of deformation in the forging. Figure 6 shows the initial and final steps of the forging process by the half sections of the cylindrical die and rectangular billet. Contours showing the levels of deformation superimposed on the final part in Figure 6(B) can be compared with the micrograph in Figure 5 to see the corresponding "X" pattern. This is more clearly seen in the color image of a quarter section of the forging shown at the beginning of the article.

By viewing a more complete sequence of images like those in the color figure, we discovered the reason for this "X" deformation pattern: the upper portion of the billet was deforming first, while the bottom half deformed less because of constraint caused by friction between the billet and the bottom of the die. We performed additional simulations using different amounts of friction between the die and the billet and found that less friction resulted in more uniform distributions of deformation. Therefore, we recommended that the die/billet interface friction be minimized. However, it is difficult in practice to carefully control interface friction during sequential forging of such a small part in closed dies. Furthermore, the analysis suggested that only a modest improvement in homogeneity would be achieved even if frictionless conditions could be obtained. The ultimate solution was to simply use a larger die and billet and then machine the final part from the central, homogeneous region of the forging.

The most significant value of this analysis was the ability to visualize the deformation process and its dependence upon die/billet interface friction. Even for this simple geometry, we must stress the utility of finite-element analysis in visualizing deformation patterns and filling of the die. For example, it was possible to examine the distribution of shear deformation at the die/billet interface graphically using the post-processor, TAURUS, to "remove" the die. This ability to view the forging process internally is not otherwise available to the manufacturing engineer and is helpful during process development. In this case, it allowed the deleterious effect of friction to be understood.

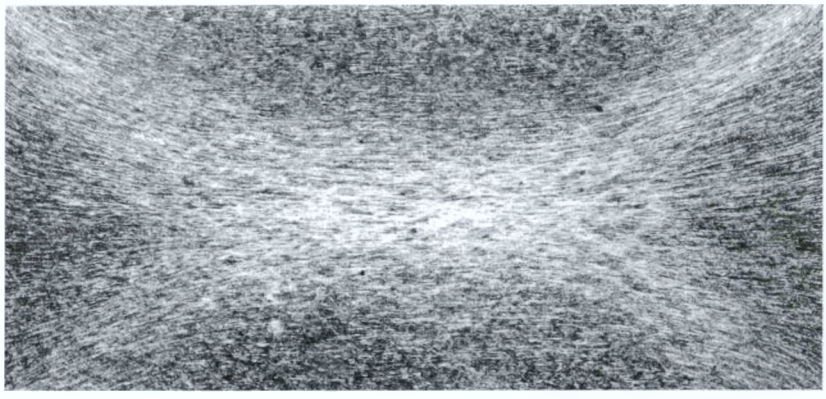


Figure 5
The contrasting light and dark regions in this specially etched half-section of a cylindrical forging show that its microstructure is inhomogeneous (see the computer-generated picture at the beginning of this article).

Box A
The Metal Forming Modeling Subgroup of the Interagency Manufacturing Operations Group

In May of 1984, a small group of engineers from Sandia National Laboratories, Lawrence Livermore National Laboratory, the Oak Ridge Y-12 facility, and the Rocky Flats Plant met informally to discuss the introduction of computer simulation of process modeling into the weapons design and manufacturing process. Participants expressed mixed optimism and concern for the feasibility of such a mammoth undertaking. From this meeting, Terry Lowe and Elane Flower organized the Interagency Metal Forming Working Group (IMFWG), a collection of experts in metal forming, finite-element analysis, metallurgy, and constitutive equation development. Championing the themes of communication, collaboration, and cooperation, the group members established working relationships. Sandia personnel conducted classes to train production plant engineers to use existing finite-element codes on realistic current process development problems. The group coordinated mechanical testing to meet the needs for materials property data in the analyses and for development of improved constitutive relations.

The IMFWG collaboration successfully provided a critical link between applied research on computational methods and constitutive modeling and the overall objective of advancing metal forming technology through process simulation. Even our very earliest analyses produced tangible benefits. Workers at the Y-12 facility received awards from the Manufacturing Engineering Society (1986) and the Department of Energy (1987) for their efforts in computer-integrated manufacturing, of which the finite element simulations were a part. The weapons community realized savings of over $4 million in the forming of metal components in the Modern Strategic Bomb, the Ground Launched Cruise Missile, the MX Missile, and the Trident II.

In December of 1986, the IMFWG changed from a grass roots collaboration of a few engineers to an official subgroup of the Interagency Manufacturing Operations Group, (IMOG). This newly christened Metal Forming Modeling subgroup of IMOG first met at Lawrence Livermore National Laboratory in April of 1987. Since that time, we have expanded our collaboration to include interaction with universities and other government laboratories. We have successfully applied finite-element simulation in numerous examples. Now one of the central challenges of the group is to better integrate computer simulation into the already established conventional design and manufacturing procedures. This challenge requires a level of coordination between design and manufacturing organizations that is beyond the scope of the technologically oriented metal forming group, but is currently being driven by increasing attention to safety, quality, and cost in weapons manufacture.

Box B
Constitutive Models for Metal Forming

Constitutive models represent the material dependent properties of solids. They are the mathematical relations that describe, for example, how rubber deforms differently from stainless steel. These relations may be written in many different forms and are derived by a variety of techniques. Different constitutive modeling approaches can be distinguished by the physical size scale of the behaviors they represent and the degree to which they are derived from physical theories. These distinctions are really parallel since models that represent a finer level of micro-structural detail tend to be more closely tied to physical theories.

At the microscopic extreme, atomic-scale models of the interactions of individual atoms are used to describe the properties of very small volumes of matter, usually comprised of less than 10,000 atoms. At the macroscopic extreme are empirical models. Though no physical size scale is implied for empirical models, they are usually based upon extensive macroscopic experimental measurements to ensure their descriptive validity, although they have no relationship to physical theories. They represent material behavior only over the range of experimental conditions from which they were derived and are strictly valid only for the specific metals, and often the specific product form (plate, rod, sheet), from which test material data were obtained. Sandia is involved in the development of models spanning the entire spectrum of physical detail. However, models with the characteristics of either extreme are not readily directed toward metal forming applications. In between these extremes, there are modeling approaches that are better suited for addressing problems in deformation processing. Two types of models lie in this domain—internal state variable models and micromechanical models.

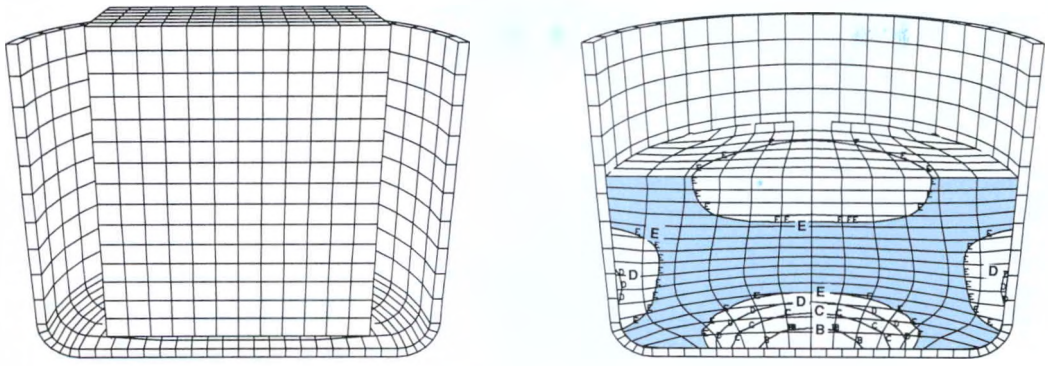
Internal-state-variable models

Internal-state-variable models are formulated in terms of a small number of variables that represent the effects of microstructure on deformation, i.e., they are physically motivated but are not necessarily linked to physical theories in a rigorous manner. These models universally show that the mechanical response is connected to the evolution of microstructure but that the details of this connection are too complex to include in a usable mathematical model. They are formulated using differing approaches and relying upon different approximations depending upon the types of deformation behavior or the particular metals of interest. For example, we have developed a constitutive model with five internal state variables that is particularly suited to represent the deformation of alloys that contain hard particles called precipitates or dispersoids. This model would have been useful in the analysis of the hemispherical forging, which happened to have been made of a metal containing precipitates.

Micromechanical models

Micromechanical models are based on detailed physical theories and, therefore, represent behaviors that internal-state-variable models cannot. Because they are computation intensive, they cannot yet be economically used in production scale, metal forming simulations. However, they can serve at least two purposes that support metal forming simulation. First they provide an analytical tool that can be used in combination with experiments to improve our understanding of microstructural processes that have been observed in metal forming. This analytical tool can then be used to represent the same processes occurring under conditions that are not easily studied experimentally, but may be of great interest in metal forming applications. For example, behaviors at the extremes of high or low temperatures, very large deformations, or very high deformation rates may be studied through micromechanical models.

To the extent that the models are successful in this first purpose, they may then be used for a second purpose—to guide the formulation of internal-state-variable models which can, in turn, be used directly in finite element analyses of metal forming. For example, we have used a micromechanical model to simulate how metal single crystals and polycrystals behave when subjected to very large deformations and then compared the results with simulations of the same deformations using internal-state-variable models. The differences between the predictions of the micromechanical model and the internal-state-variable models tell us what we need to do to make the internal-state-variable models better.



(A) (B)

Figure 6. These half-section views of the finite-element meshes show the (A) initial and (B) final steps of a rectangular billet being pressed into a cylindrical die. Contours showing the gradient in effective plastic strain indicate the type of inhomogeneity.

Even though we have demonstrated that finite-element simulation is a valuable tool for developing manufacturing processes, we have not forgotten the need to assemble the other ingredients required to use this tool in a manufacturing environment. Our simulations were performed using the resources of a national laboratory.

Can finite-element simulations provide the same benefits to manufacturing engineers with more limited resources? The answer is yes because of the efforts of a special group of engineers and scientists from different specialties who joined to make the whole effort work (see Box A). Sandia's contribution to metal forming technology currently focuses on improving the capabilities of finite-element codes for simulating metal forming and on formulating better mathematical descriptions of the intrinsic properties of metals. The latter contribution is particularly important because our ability to predict the effects of deformation and heating on the mechanical properties of forgings depends entirely on how well we can represent changes in the intrinsic properties of metal while it is being deformed. Fortunately, Sandia has experts who specialize in formulating models of how deformation affects metals (see Box B).

Computer simulation using finite-element codes has proved to be effective in the manufacture of forged metal components.

Our research at Sandia has shown that computer simulations using finite-element codes can aid the manufacture of metal components by forging. As we continue our research to expand this capability, we must also consider the changes that the new technology will promote. We can look forward to a time when the entire process from design to manufacture can be integrated using computers.

Computer-aided design and drafting is already commonplace, as is the use of computers for controlling manufacturing operations such as machining. Now that we have shown that computer-aided simulation can also guide the intermediate steps of developing manufacturing processes such as forging, it is reasonable to anticipate that computer-aided design, simulation, and manufacturing can all be combined in computer-integrated manufacturing. Such an integrated approach is already being realized for simple design-to-manufacture sequences, but is just becoming possible for metal forming. It is now conceivable that a small group of individuals in a single room with a computer can serve nearly the same function that in the past has been performed by a much larger cast of supervisors, engineers, and technicians working in separate organizations, or within DOE's Nuclear Weapons Complex, or in separate companies.

The concept of integration is attractive but challenges us to discard organizational barriers and adopt new notions of product design and manufacturing. This challenge is no less than that of developing the computer simulation technology in the first place, but really is just another element of our complex modern society that distinguishes us from early civilized man.

For more information, please call ***Terry Lowe,*** (415) 294-3187.

Characterization and Modeling of Weld Microsegregation

We studied the microstructure of welds made with different processes to improve their reliability.

In the history of materials fabrication and applications, welding is the preferred method of joining parts in many structures and assemblies because it can provide rigid, mechanically strong, reliable connections. Our experience shows that welded joints formed under proper conditions remain reliable after many years of service. Often welding applications in weapon systems involve unique combinations of materials and complicated shapes seldom encountered in conventional welding. To select the kinds of welds that will be the most reliable and easiest to fabricate for these applications, we must understand the various processes that can be used.

Our research at Sandia focuses on the fabrication of high-pressure, gas-storage vessels containing corrosive gases that must retain their structural integrity for long times under weapon stockpile conditions. To improve the welds for these vessels, we have investigated the process of solidification thoroughly at the microstructural level to explain features of crystallization that are observable only under high magnification (Figure 1). This research also applies to improving the quality of castings since the basic solidification processes in castings are similar to those of welding.

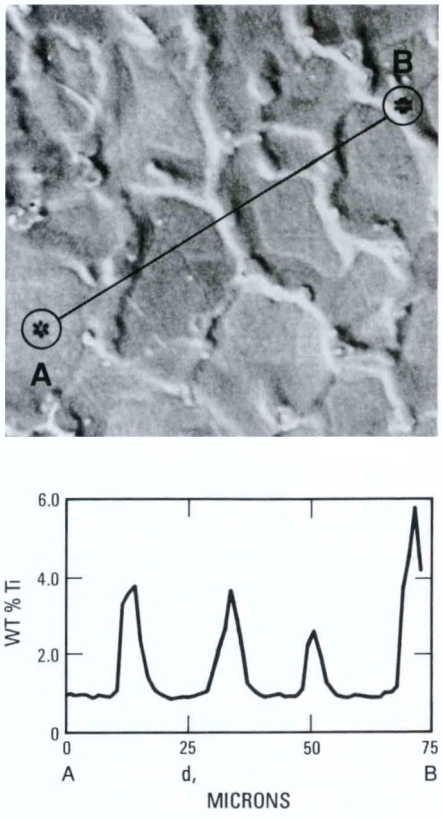


Figure 1.
This scanning electron microscope (SEM) photograph shows evidence of weld microsegregation occurring in a stainless steel alloy. The cell cores shown in this transverse section solidify first. As solidification progresses, material composition changes. Microprobe analysis shows periodic variations in titanium, one of the main strengthening elements across the cellular structure.

Compositional variations within welds reduce their strengths.

Solidification in most two or more component systems—for example, salt and water—occurs over a temperature range. The composition of both the solid and liquid—in the above example, the salt content of the liquid and the ice—continually changes as solidification proceeds. Similarly, in a metal alloy, as the temperature of a solidifying mixture of liquid and solid decreases, variations in composition (called microsegregation) are frozen into the solid crystals. Since the thermodynamic equilibrium concentration of the alloying elements changes as a function of temperature, the first solid to form will have a different composition from that of the last (Box A). At the same time, near completion of

Box A
Understanding Solidification

Solidification is described by equilibrium phase diagrams. Aluminum-copper alloys are a particularly simple example of solidification because the liquidus and solidus phase lines are straight lines on the phase diagram (Figure A–1). At temperatures and compositions within the liquid-plus-solid two-phase field, the composition of the liquid and solid in equilibrium at any given temperature corresponds to the compositions of the liquidus and solidus, respectively. Thus, a given alloy composition begins to solidify when it reaches its corresponding liquidus temperature and is completely solidified at its corresponding solidus temperature when cooled under equilibrium conditions. Figure 1 shows that the addition of copper both lowers the solidification temperature of pure aluminum and results in solidification occurring over a temperature range. The equilibrium cooling path for an aluminum-copper alloy is a vertical line on the diagram through its average composition; Figure A–1 shows a 3% copper alloy that starts to solidify with a solid composition of 0.5% copper at approximately 650°C. This would correspond to the composition of the tip of a solidifying dendrite in a cell. As the temperature drops, more and more solid forms. If equilibrium exists, as cooling continues, copper diffuses from the liquid into the solid, maintaining a uniform solid composition which increases in copper until the material homogeneously solidifies at about 590°C. Below this temperature, under equilibrium conditions, no remaining liquid exists. Such equilibrium solidification is nearly impossible to achieve, however, and composition homogeneity generally does not occur.

Microsegregation refers to nonuniformity of composition throughout individual crystals which occurs because of nonequilibrium solidification. In the above example, if solid-state diffusion is limited, as the weld cools, the composition of the first melt to solidify will remain near 0.5% in the 3% copper alloy. As the temperature decreases, the composition of both the solidifying solid and remaining liquid increases in copper because of the reduced amount of Cu in the original solid. As a result, the last liquid will contain copper at the eutectic composition of 33%, at which point it will solidify into two phases: 5.6% Cu and $CuAl_2$. Thus, the solid formed is very nonuniform in composition.

While this nonequilibrium solidification promotes segregation, solid-state diffusion of the components reduces it. In some alloys diffusion may be sufficient during cooling to neutralize segregation. More often, however, the solid must be reheated in a heat-treatment cycle and held at elevated temperature for a prescribed time to allow diffusion to alter solidification structures.

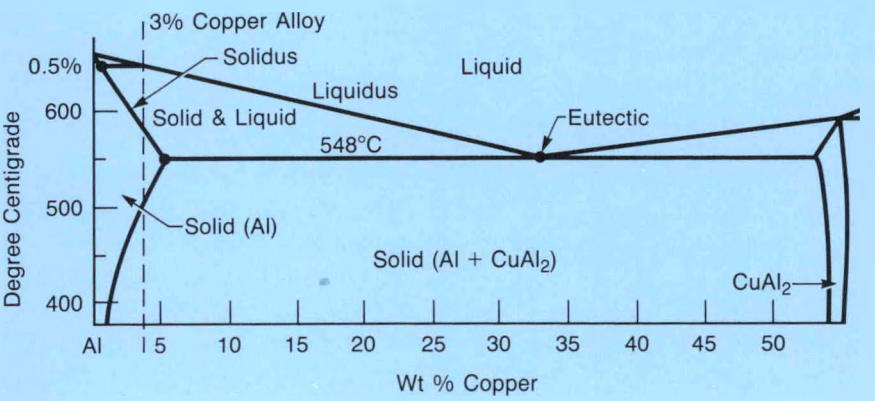


Figure A–1.
Aluminum-Copper Phase Diagram

solidification, nonequilibrium segregation of low-melting liquids, along with internal stresses and certain microstructure features, may cause cracking, a reduction of the joint strength, and a degradation of other properties.

A primary goal of our research was to understand the differences in weld structures which can affect weld-cracking susceptibilities that are caused by variations in alloy compositions and cooling rates. Cooling rates during welding can vary from about 100°C/sec (conventional gas-tungsten arc, GTA, welding) to 100,000°C/sec (high-energy-density electron-beam and laser welding).

The rapid cooling rates of high-energy-density welds are a consequence of highly localized heat source melting material with little heat being deposited in the surrounding base metal. During solidification, the heat contained in the weld fusion zone is quickly cooled by conduction into the surrounding cooler base material. Welding speeds can be very high, ~100 in/min; or pulsed laser welds can be made over short durations (~1 millisecond). Thus, while the energy density is very high, the total heat input is very low.

On the other hand, with GTA welding, the heat source is more widely distributed; the welding process is much slower (5 to 10 in/min); and the total heat input is much larger. Consequently, the cooling rates are also much slower than those of the high energy density welds.

To better understand microsegregation, or variations in structure caused by cooling rate differences, we compared weld microstructures from different types of processes with predictions of theoretical solidification models. The following paragraphs summarize these findings.

Different welding processes have different cooling rates and alter the amount of microsegregation in a weld.

Because of the complexity of the weld-solidification process, we selected well defined, two-component systems, including aluminum-copper, for our first studies. Aluminum-copper alloys have close-packed, face-centered, cubic crystal structures with low diffusion rates that tend to promote microsegregation.

In more open structures—such as body-centered cubic crystals—diffusion rates are generally faster and some homogenizing of the solid can occur during solidification and cooling. Our results showed that microsegregation in aluminum-copper welds depends strongly on the rate of cooling, but in a manner entirely different from what we had expected (Figures 2 and 3).

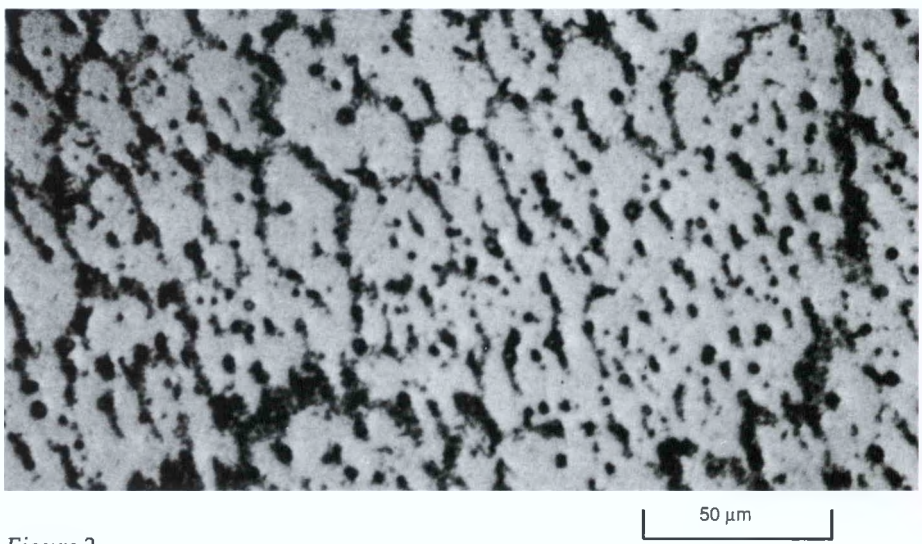
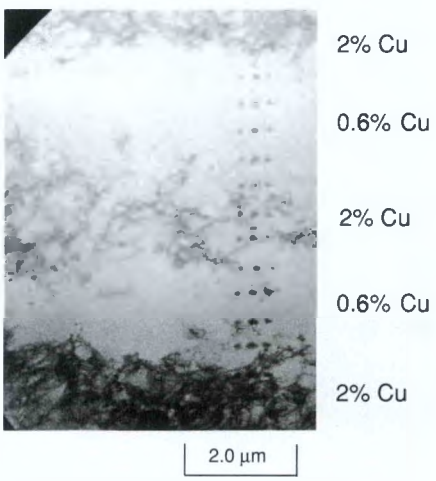


Figure 2.
This photograph shows microsegregation in a typical gas-tungsten arc (GTA) weld of an aluminum-2.0% copper alloy. Alloys of this composition are "decorated" by heat treating them after welding to precipitate second phases at the copper-rich cell boundaries. We find more copper than average (about 5 to 6%) in the vicinity of the cell boundaries, and less copper than average (about 0.8%) in the middle of the cells, the core region. This shows that diffusion processes have not had time to make the composition uniform throughout the alloy.

Figure 3.
We also see microsegregation in the high-energy-density, rapidly solidifying aluminum-1.1% copper welds. For chemical analysis, special high-resolution, analytical electron microscopy techniques were required because of the small cell size (about 2 µm). Dark spots in groups of three show regions analyzed.



Although high cooling rates were expected to increase microsegregation (because less time is available to approach uniformity of composition by diffusion), we found instead that high energy density welds with faster cooling rates were more homogeneous than welds that solidified at low cooling rates. We attributed these unexpected differences to surface tension and dendrite tip curvature effects when finer crystal structures are formed during the rapid cooling.[1] As the solidification velocity increases, undercooling increases; that is, solidification starts at a lower temperature at a composition closer to that of the bulk alloy. Our experimental results agreed with predictions based on our finite-differences model (Box B) and showed that solid state diffusion—unlike undercooling—had little effect in reducing

[1]Crystals, formed during solidification, can form as a hard core structure with branches resembling a tree and are called cells or dendrites.

Box B
A Computer-Based Model for Analyzing Microsegregation

Using finite-difference analysis in a cylindrical geometry, we are developing mathematical solidification models to assist us in interpreting weld compositional changes and microstructures. Cellular structures formed during solidification are approximated as needles sticking into the melt and expanding in diameter until solidification is complete. We assume equilibrium conditions at the growing solid-liquid interface and use diffusion coefficients that depend only on temperature. We predict compositional changes for a given alloy by assuming a constant cooling rate and calculating the composition and amount of solid deposited at the dendrite outer surfaces during each time increment. Solid state diffusion is also coupled into the calculation at each time step. We modified an existing code, DIFFUSE, for our requirements, changed the code boundary conditions, added phase diagram information, and incorporated a moving solid-liquid interface option.

For example, in the application of our mathematical model, we simulated solidification of an iron-21% chromium, 14% nickel-alloy at a cooling rate of 300°C/sec (Figure B–1). There are two possibilities: the liquid was allowed to solidify to 100% body-centered cubic ferrite or to 100% face-centered cubic austenite (refer to Box C). For the latter, very little diffusion occurred, and the final structure is highly segregated with up to 22.5% nickel at the cell boundaries and only 11.8% nickel at the center of the cells. Because of rapid diffusion for ferrite solidification, we observed a much more uniform structure with 14.1% nickel at the boundaries and 13.3% nickel at the centers of the cells. Chromium concentrations behaved in a similar manner. Actual welds in this alloy composition, solidified as austenite, were found to be highly segregated and very similar to the composition we predicted. These studies show that segregation is due to the very limited solid-state diffusion occurring in the austenite, rather than the particular characteristics of the phase diagram.

We compared our results with real welds to determine which solidification path is more representative of actual cooling rates. Some precipitation-hardened alloys are quite homogeneous under these welding conditions, supporting the ferrite path; other alloys are highly segregated, supporting the austenite path. Mathematical simulation of weld solidification in this manner is new to the welding community and offers us the opportunity to unravel very complex solidification and solid-state transformation histories.

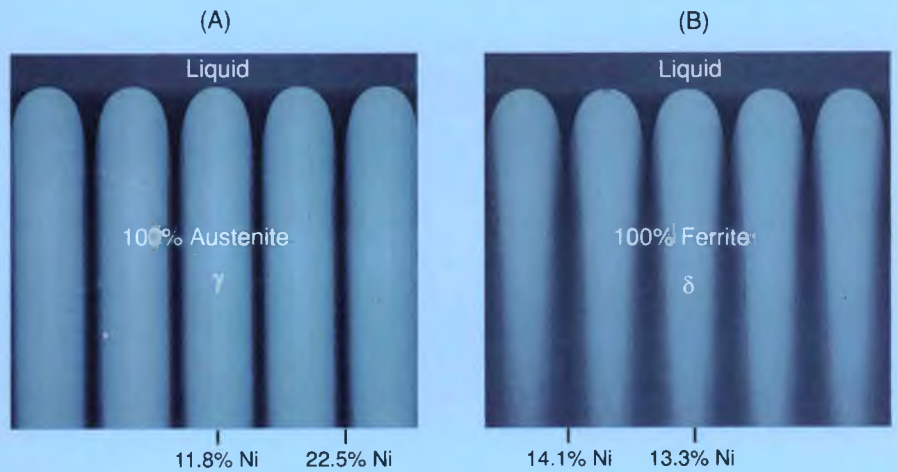


Figure B–1.
Computer simulation shows solidification of an iron-21% chromium, 14% nickel-alloy weld at a cooling rate of 300°C/s. Color intensity corresponds to Ni concentration. Photograph (A) shows concentration profiles when the material is assumed to solidify as face-centered cubic austenite. Photograph (B) shows the concentration profiles predicted when solidification is assumed to be a body-centered cubic ferrite structure.

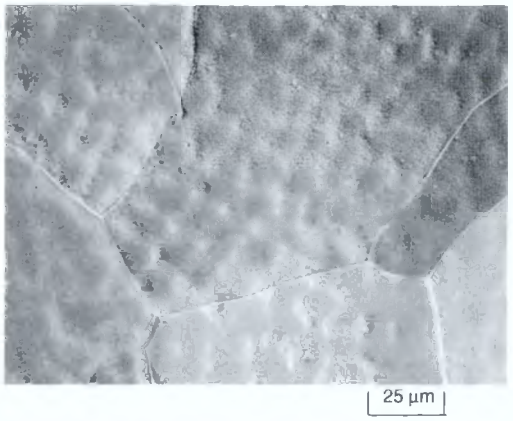


Figure 4. Microsegregation structures of iron-2% niobium alloy gas-tungsten arc (GTA) welds are much smaller compared to aluminum-copper welds because of diffusion. The initial as solidified cell concentration of 0.5% Nb has increased to 1.8% as a result of solid-state diffusion.

microsegregation even though the cell size and diffusion distances decrease with increased cooling rates. We concluded from these studies that an increased cooling rate can actually be beneficial for aluminum-copper alloys because it reduces microsegregation.

In contrast to the strong presence of microsegregation in aluminum-copper welds, welds of iron containing 2% niobium appeared remarkably homogeneous (Figure 4). Our mathematical model showed that the absence of microsegregation in iron-niobium systems was due to the rapid diffusion of niobium during cooling. This observation is consistent with the open nature of the body-centered cubic crystal structure of the alloy, which enhances diffusion. In the high energy density (fast cooling) welds, diffusion times are very short, but a high degree of homogenization still occurs because of the smaller cell sizes and the shorter required diffusion distances. The degree of homogenization in the weld was much greater than would have resulted from solidification undercooling alone.

Our results have helped us understand the properties in engineering alloy systems.

From our results with simple systems such as aluminum-copper and iron-niobium alloys, we have been able to understand and, in some cases, control the microstructure and resulting properties of more complicated engineering alloy systems used in weapons. We know, for example, that weld solidification and cooling of iron-nickel-chromium alloys, the basis of many engineering stainless steel alloys such as 18-8, produce a number of different microstructures in the solidified state (Box C) and Figure 5. Also, our results from welds of three different iron-nickel-chromium compositions with chromium/nickel ratios of 1.5, 1.73, and 1.92 demonstrate how welds are affected by microsegregation. To determine the effect of cooling rate on microsegregation and structure, we formed electron beam and laser welds over the top of GTA welds. Figure 6 shows these results. The cellular solidification structure is readily revealed by metallographic preparation and acid etching that attack the varying alloy compositions at different rates.

Welds of the alloy with the lowest chromium/nickel ratio (65% iron, 14% nickel, and 21% chromium) provide a classic example of microsegregation during welding (Figure 7). In this alloy system, solidification occurs over a large temperature range, and small amounts of liquid are

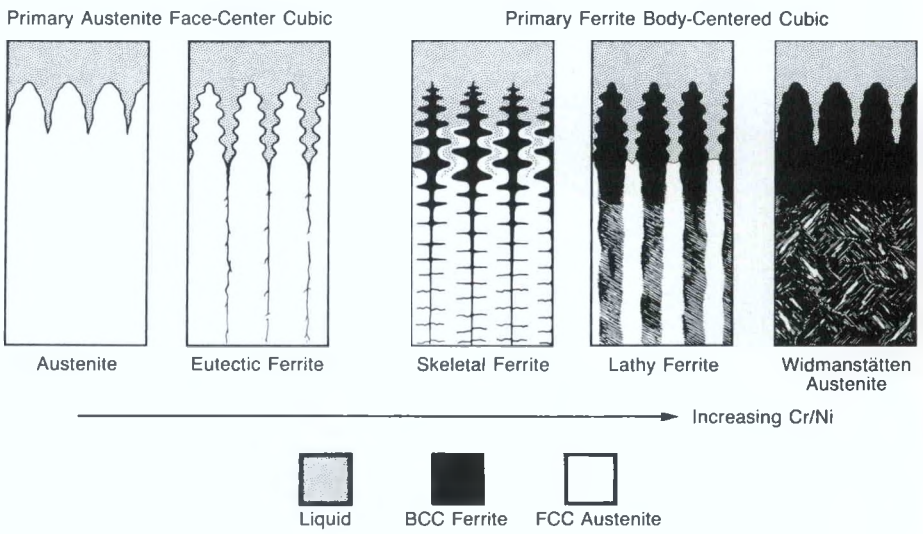


Figure 5.
The competition of microsegregation processes and diffusion can produce a number of different weld morphologies for the solidified states of iron-nickel-chromium alloys as illustrated in these sketches. Changes in the liquid composition during solidification can result in a change in solidification structure. Also a solid state transformation during cooling of the body-centered cubic material complicates the weld microstructure.

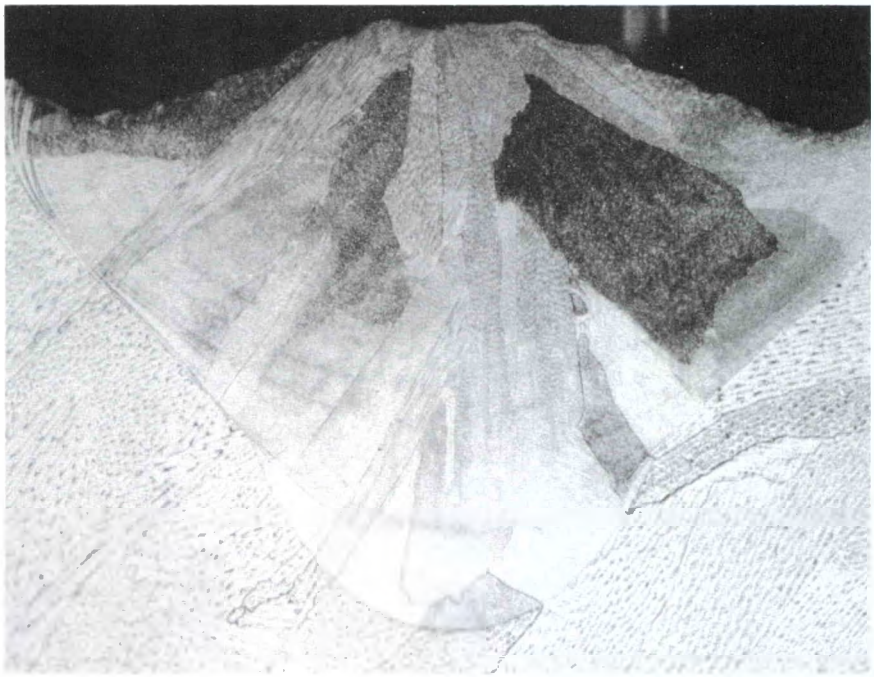
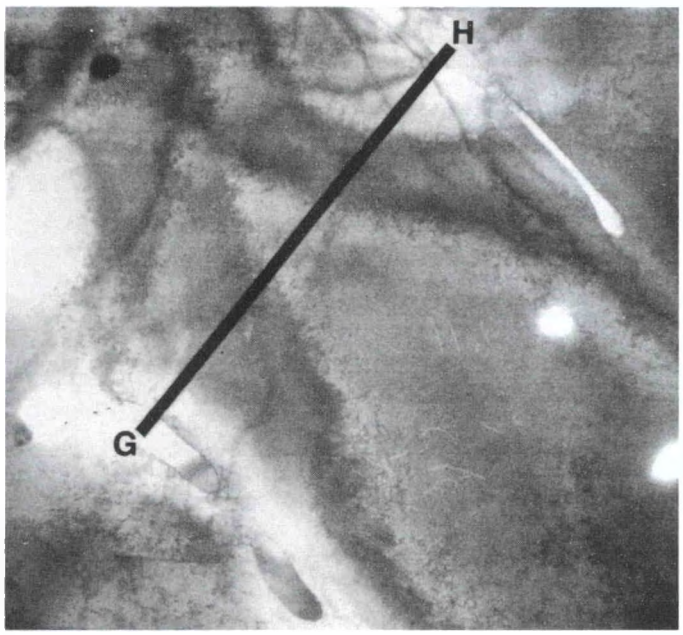
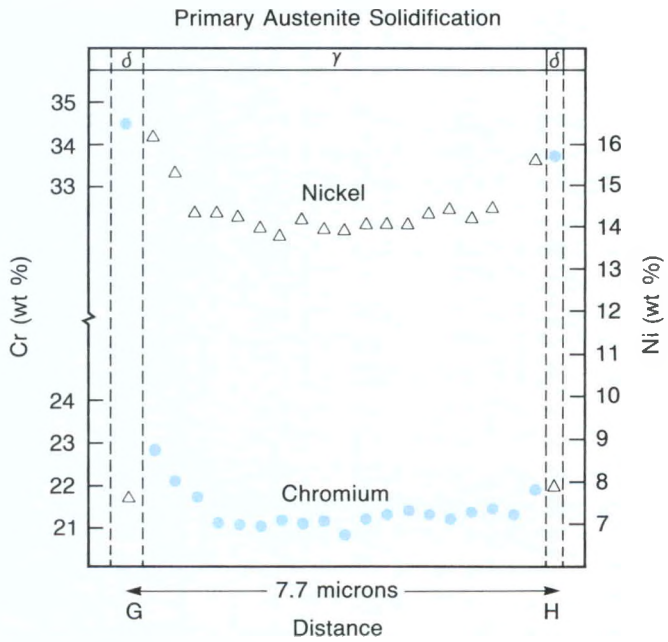


Figure 6.
The influence of different cooling rates on weld solidification microstructure can be seen in this photograph of a laser weld on top of a GTA weld. We observe a much smaller cell size in high-energy-density welds that rapidly solidify. Ferrite is observed in the GTA cell boundaries, but not in the laser weld boundaries, because of reduced microsegregation resulting from dendrite-tip undercooling.





Primary Austenite Solidification
STEM Analysis

Figure 7.
Microsegregation in a 65% iron, 14% nickel, and 21% chromium alloy solidifying as face-centered cubic austenite when GTA welded is evident in analytical electron microscopy results. Both nickel and chromium compositions are low in the core but increase to a maximum in the last material to solidify, resulting in changes in solidification structure. (This is possible in a three-component system.)

Figure 8.
Gas-tungsten arc welds of 70% iron, 11% nickel and 19% chromium solidify as BCC ferrite, followed by FCC austenite, leaving some residual ferrite along the solidification cell cores (see Figure 5). The laser weld has the characteristic fine cellular structure resulting from austenite solidification; in the high-energy welds, we have a change in the primary phase of solidification. The undercooling occurring during rapid solidification results in a change in composition of the solidifying solid that favors austenite solidification.

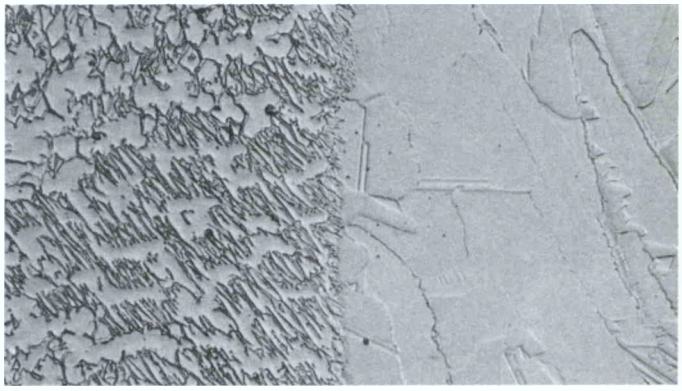


Figure 9.
This weld shows the highest chromium/nickel content alloy of 65% iron, 12% nickel, and 23% chromium. The laser weld, all austenite, does not exhibit a cellular structure characteristic of austenite solidification, but has, instead, a fairly fine grain structure with jagged boundaries. This structure results from ferrite solidification, complete homogenization, and then a diffusionless solid-state transformation of ferrite to austenite.

present along the solid cell boundaries during the last stages of solidification. If significant stresses develop before the final stages of solidification are complete, weld cracking can result.

In contrast, GTA welds of the alloy with an intermediate chromium/nickel ratio of 1.73, only slightly different in composition (70% iron, 11% nickel and 19% chromium), showed a GTA weld morphology with greatly reduced microsegregation because this material solidified as body-centered cubic ferrite (Figure 8). However, the electron beam weld on the same structure was found to solidify as face-centered cubic austenite.

Understanding these solidification processes enabled us to evaluate the consequences of changing, for example, from a GTA process that is known to produce a structurally sound weld to a higher-energy density welding process. The solidification mode of the new process can be different, and this difference may cause the weld to crack.

In further experiments, we found that GTA welds of 65% iron, 12% nickel, and 23% chromium (with the highest chromium/nickel ratio of 1.92) solidified differently than did the low- and medium-ratio alloys (Figure 9). The GTA weld solidified almost completely as ferrite but transformed mainly to austenite by a diffusion-controlled transformation when cooled through the two-phase ferrite and austenite phase field shown in Figure C–1. The laser weld morphology of this composition was quite different from that of GTA welds and was single phase with almost uniform cell composition.

Results from our computer simulation confirmed that even though segregation existed at one time during solidification—as body-centered cubic material—an almost homogeneous structure should exist. However, during rapid cooling of the laser weld, the diffusion-controlled transformation of ferrite to austenite was suppressed, and instead, the laser weld transformed directly to austenite via a diffusionless transformation. This structure, in which microsegregation was greatly reduced, is very resistant to weld cracking.

Thus, our studies of weld microsegregation in iron-nickel-chromium systems showed that the quality of a weld depends not only on rather small variations in alloy content but also on the rates of solidification associated with various welding processes. High-energy-density welds with faster cooling rates may result in either highly segregated or homogeneous structures, depending upon alloy composition and the degree of weld undercooling.

Box C
The Structure of Welds in Iron-Nickel-Chromium Alloys

Many of the commonly used stainless steel alloys undergo complex changes as they solidify. This system has two solid phases, ferrite (body-centered cubic crystal structure) and austenite (face-centered cubic crystal structure), with a number of possible combinations. As a result of minor compositional variations, one lot of the material can solidify as austenite; another can solidify as ferrite; or as will be shown, one phase of the solidification can be followed by the other (see Figure 5).

The solidification mode that occurs is important to us because welds in which ferrite is the first phase to solidify are resistant to cracking. We expect ferrite to exhibit high diffusion rates that reduce microsegregation because its body-centered cubic structure is the same as the iron-niobium alloy that we previously examined. Austenite, a face-centered cubic structure similar to the aluminum-copper alloy, will have slower diffusion rates and, therefore, more microsegregation.

Our studies included an iron-nickel-chromium composition of about 14% nickel and 21% chromium to illustrate our observed reactions. Cooling this alloy, shown along line A-A' in Figure C–1, begins with solidification of austenite, but some ferrite also begins to form when the composition of the liquid intersects that of the eutectic. (Since this is a three-component system, the liquidus and solidus compositions shown do not necessarily correspond to those in equilibrium at the given temperature.) Upon complete solidification, the alloy is a mixture of the two. As the temperature drops further, the ferrite phase transforms under equilibrium conditions to 100% austenite. Complete transformation of ferrite to austenite may not occur in welds; however, because of limited time for diffusion, metal at the cell boundaries remains overly rich in chromium, retaining the ferrite phase. With higher nickel contents, the melt can also solidify as 100% austenite before the last liquid reaches that of the eutectic.

Other sequences of transformations are observed on the chromium rich side. For example, alloys of 12% nickel and 23% chromium would cool as shown along line B-B' in Figure C–1. This liquid composition may solidify as ferrite followed by austenite solidification if sufficient microsegregation occurs. However, since ferrite diffuses faster, microsegregation can be reduced and the material may solidify completely as ferrite. As it cools, the material passes through the two-phase field of austenite and ferrite. With lower cooling rates of GTA welds, a large fraction of ferrite transforms to austenite (Figure 5). However, with high cooling rates of laser and electron beam welds, this diffusion-controlled transformation can be suppressed and the material can change through a diffusionless transformation directly to austenite, as can be seen in Figure 9.

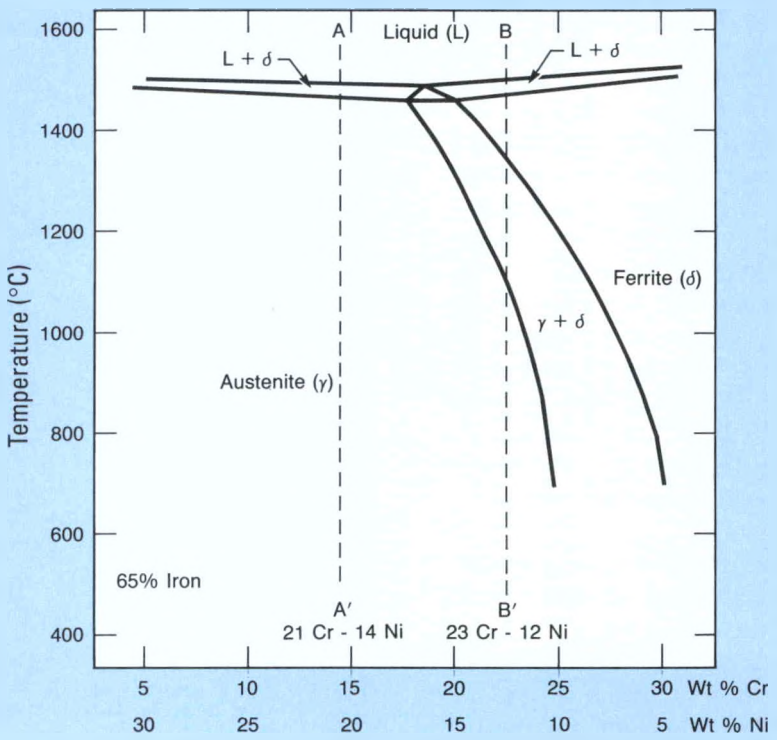


Figure C-1.
This graph illustrates a constant Fe vertical section of the Fe-Ni-Cr diagram showing phase fields intersected during weld solidification and cooling for alloys that solidified initially as austenite and finite.

Finally, simulation of weld solidification conditions with a mathematical model demonstrated that diffusion of elements can play an important role in reducing microsegregation.

Knowledge of microsegregation helps produce more reliable welds in weapon components.

As a result of our research, we can identify the best weld material composition and welding process for each weld in weapon components. In these designs, reliable welds are a particular concern because limitations on size and weight require designing to the optimum strengths of all materials. This means that weld defects must be extremely small or completely absent. In addition, high performance alloys used in weapon systems often require high alloy additions, making them more susceptible to defect formation during welding.

The high-strength, stainless steel plunger assembly (Figure 10) is an example of the complex shapes that we routinely encounter. This part must withstand internal pressures of 2000 psi hydrogen at temperatures as low as –30°C without any degradation of strength. To meet these requirements, we found that the plunger assembly could be fabricated best with three fusion welding techniques. GTA welds were used on the body, and laser and electron-beam welds were used for temperature-sensitive tube and plunger components.

In designing the welds, experimental data from our study of the iron-nickel-chromium alloys suggested that lot-to-lot variations in the compositions of materials used for the electron-beam and laser welded components would cause them to be susceptible to cracking. Therefore, we specified alloy compositions that we were confident would solidify as body-centered cubic ferrite. Furthermore, for optimum structural performance in hydrogen, we limited the final amount of ferrite in the base material and weld structure.

Extensive testing of this component has shown that our welds meet all design specifications. Greater weld reliability, achieved by controlling the compositions of the starting materials, has resulted in a decrease in weld microsegregation and the elimination of cracking. Encouraged by these results, we plan to extend our studies to other complex alloy systems so that we can continue to expand our knowledge of the influence of solidification and solid-state transformations on weld strength. This will result in better performance and more reliable systems at reduced costs because of fewer weld rejects in complex weapon systems that rely heavily on welding as a critical fabrication sequence. Our findings in this important area will be extremely valuable in future decisions on which welds will be most effective in our component designs.

For more information, please call ***John Brooks***, (415) 294-2051, or ***Michael Baskes***, (415) 294-3226.

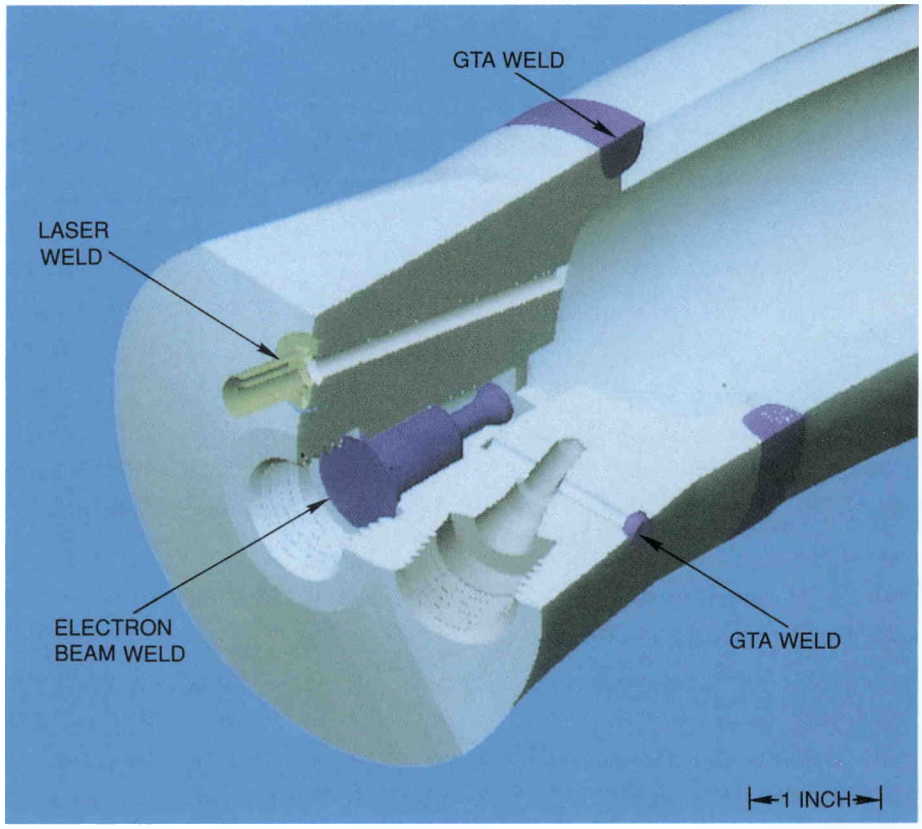


Figure 10.
This stainless steel plunger assembly required three different types of welds. To minimize weld-cracking susceptibility, the materials were tightly specified to those compositions that result in a body-centered cubic ferrite structure upon solidification after welding.