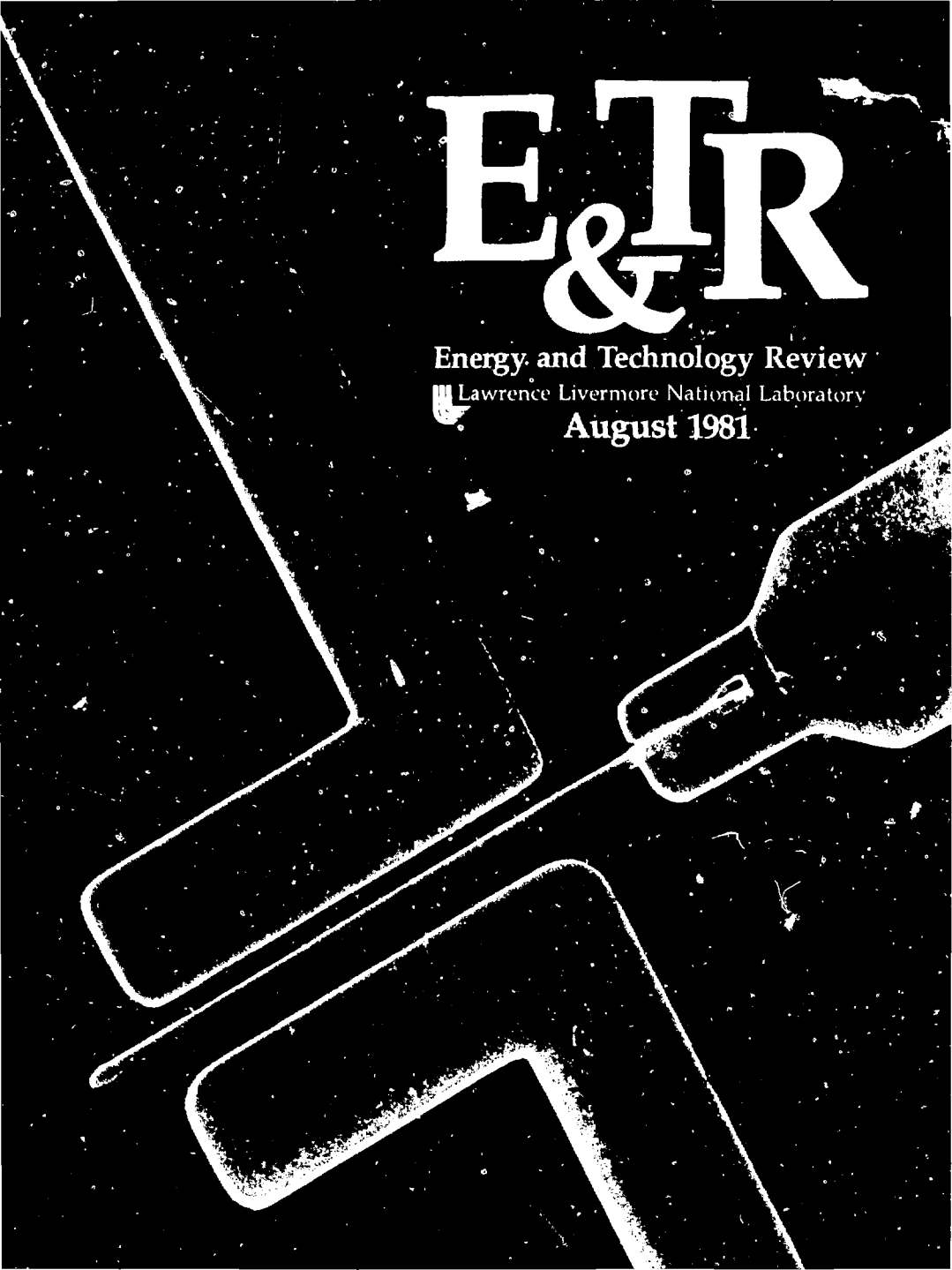

E&TR
Energy and Technology Review
Lawrence Livermore National Laboratory
August 1981

E&TR

Energy and Technology Review

Lawrence Livermore National Laboratory

August 1981

Brief ii
Revised Hiroshima and Nagasaki Dose Estimates

Predicting Integrated-Circuit Performance 1
Our new gallium arsenide integrated circuits may be made [illegible] faster than present silicon circuits

DYNA3D: Modeling Material Deformation 12
DYNA3D is a computer code that models hydrodynamics and [illegible] mechanics in three dimensions for studies of the interaction and deformation of various materials under stress

Oil and Gas Resources: Predicting the Unpredictable 22
Long range forecasts of oil and gas supplies are vital planning tools that are difficult to make even when prices are stable and market forces are the dominant driving mechanism. They become exceedingly uncertain when state-run national and political decisions are the controlling factors

Brief

Revised Hiroshima and Nagasaki Dose Estimates

Reliable estimates of the radiation doses received by survivors of the atomic bomb explosions at Hiroshima and Nagasaki are important to our understanding of the effects of radiation on human beings. Radiobiologists have correlated these dose estimates with the subsequent health history of the survivors to assess the long-term effects of exposure and the relative importance of neutrons and gamma rays. These correlations have been vital factors in establishing worldwide standards of radiation safety that are applicable to the commercial nuclear industry as well as to other activities involving sources of radiation.

We have made new estimates of the neutron and gamma-ray doses at Hiroshima and Nagasaki.[1] They result from (1) our use of the current Los Alamos National Laboratory best estimate of the bomb outputs; (2) our calculations of neutron and gamma-ray transport through the atmosphere (based on validated air-over-ground calculational tools, validated air cross sections, and improved estimates of atmospheric conditions); and (3) a comprehensive evaluation of existing test data on delayed gamma rays. Our new dose estimates, shown in Fig. 1, have important implications for radiobiological studies based on Hiroshima and Nagasaki data.

The currently accepted estimates of dose are known as T65D (tentative 1965 dose) and were prepared largely by Oak Ridge National Laboratory.[2] The T65D estimates were supported by interpretations of measurements made by Japanese laboratories of cobalt activation in steel reinforcing bars in concrete buildings at Hiroshima and Nagasaki. Any revision of the T65D estimates should be consistent with these cobalt activation data.

The largest difference between our estimates and the T65D values occurs in the neutron-dose estimate for Hiroshima. To subject this difference to the most critical examination, we have compared our neutron-dose estimate with the most distant (i.e., the most sensitive) activation data available, taken 1180 m from ground zero.

We find that our neutron activation estimates are in good agreement with the

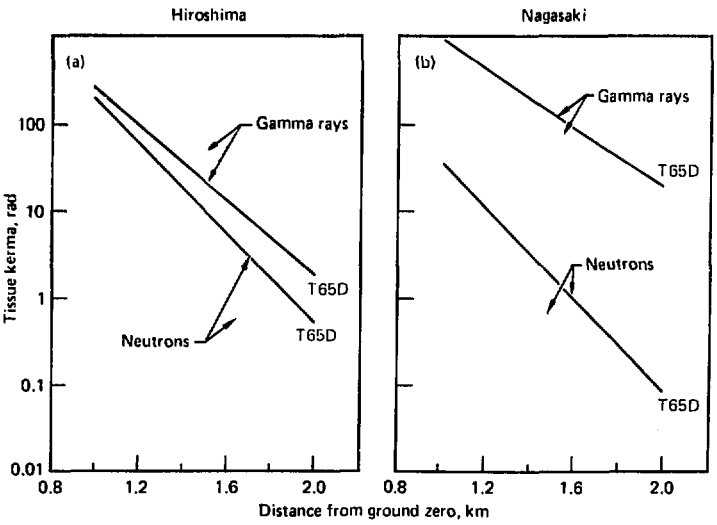


Fig. 1

LLNL estimates of free-in-air radiation doses at (a) Hiroshima and (b) Nagasaki. Earlier dose figures, known as T65D (tentative 1965 dose) estimates, are also shown. Our revised dose estimates are based on (1) our use of the current Los Alamos National Laboratory best estimates of the bomb output; (2) our calculations of neutron and gamma-ray transport through the atmosphere (based on validated air-over-ground calculational tools, validated air cross sections, and improved estimates of atmospheric conditions); and (3) a comprehensive evaluation of existing test data on delayed gamma rays. The most remarkable change in the dosimetry is that now the neutron dose at Hiroshima represents a fraction of the total dose (neutrons plus gamma rays) that is much closer to the fraction at Nagasaki. Gamma radiation greatly exceeds neutrons at both cities.

measured cobalt activation data, giving added credibility to the Los Alamos spectrum calculations as well as to our radiation transport calculations. We also tested our procedures by performing our calculations using the To5D assumptions about neutron spectra in place of the modern Los Alamos estimate. With this approach, we were able to duplicate both the To5D and cobalt-based values of neutron dose. The results have given us increased confidence that our neutron calculations are correct and that the major difference between our work and previous work lies in the difference in assumptions.

An important study published by Rossi and Mays in 1978 related leukemia mortality at Hiroshima and Nagasaki to To5D doses.[3] The results, summarized in Fig. 2a, show a clear difference in leukemia risk between Hiroshima and Nagasaki, from which these investigators inferred high values of dose-dependent neutron relative biological effectiveness (RBE).

We repeated the procedure used by Rossi and Mays but replaced the To5D-based free-in-air dose data with our own values (from Fig. 1). Our results (Fig. 2b) show no difference between the Hiroshima and Nagasaki curves. Therefore, leukemia experience at Hiroshima and Nagasaki no longer invites an inference of neutron RBEs. Similar analyses have been done by the Biomedical Sciences Division for other endpoints.[4] (In contrast to leukemia, chromosome aberrations and total cancer mortality continue to show significant intercity differences that may reflect elevated neutron RBEs.)

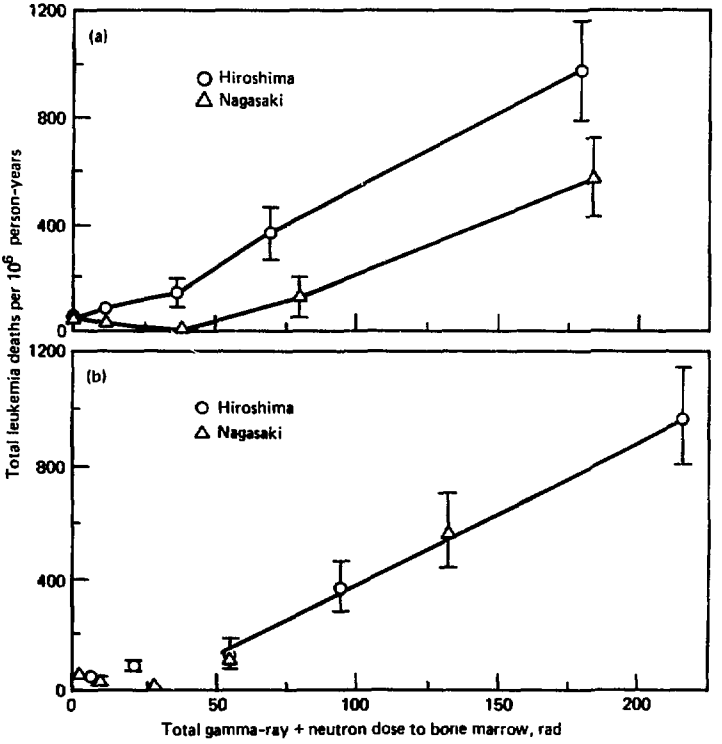


Fig. 2

(a) Results of a 1978 study by Rossi and Mays on leukemia deaths at Hiroshima and Nagasaki. They used available data that relate leukemia mortality among bomb survivors to To5D doses in rads of tissue kerma adjusted for attenuation due to structural shielding and the body's self-shielding. Their results indicate that leukemia mortality as a function of total (gamma-ray plus neutron) dose to the bone marrow varies significantly between the two cities. They postulated that this difference must have been caused by the much higher neutron component in the To5D Hiroshima dose (compared to that for Nagasaki). Accordingly, to reconcile the difference, they deduced very high values of dose-dependent relative biological effectiveness (RBE) for neutrons. (b) We have repeated the procedure used by Rossi and Mays but have replaced the To5D-based free-in-air dose data with our own values (from Fig. 1), adjusted (as a first approximation) in accordance with the same structural and body shielding factors used by Rossi and Mays. The effect of substituting our doses is to eliminate the differences between the Hiroshima and Nagasaki curves and thereby to eliminate the rationale for inferring neutron RBEs much larger than one from these data. (The error bars on leukemia mortality represent one standard deviation based on the sample size. The uncertainties for dose values are not shown on this figure.)

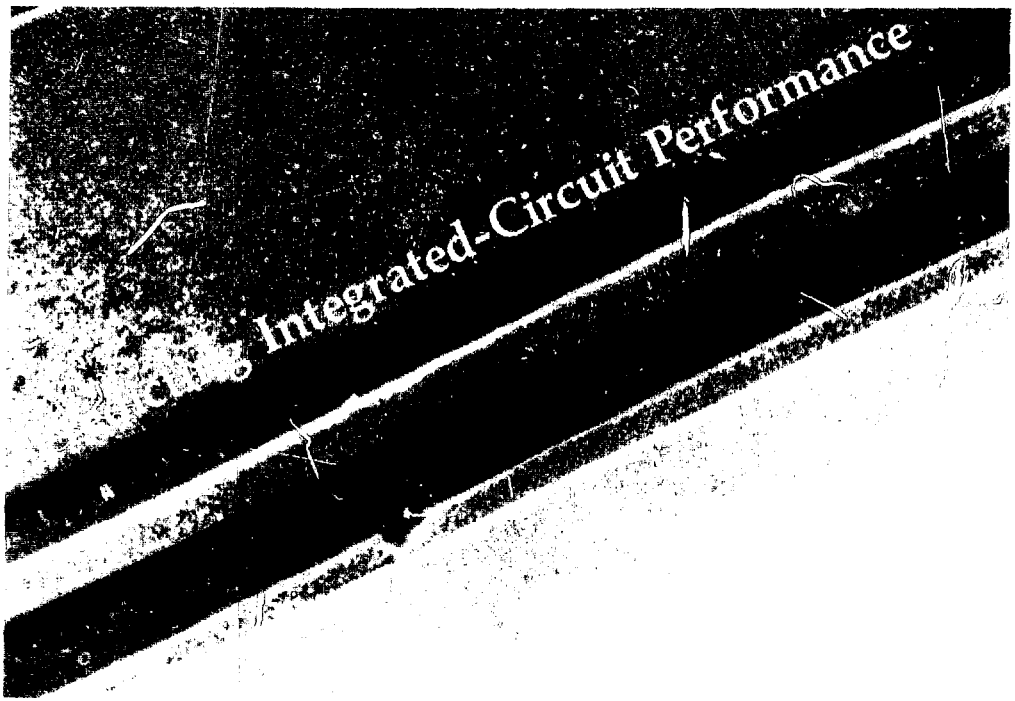
Some adjustments to our dose estimates shown in Fig. 1 may occur in the future as a result of continued investigation and further evaluation, but, in our judgment, such adjustments will not result in a change in the neutron dose of more than a factor of two at any interesting range, and the change probably will be less than that. Gamma-ray doses are subject to less uncertainty than are neutron doses. Additional effort is appropriate to reduce these uncertainties in our work and also to re-evaluate the conversion from our free-in-air doses to marrow doses.

To summarize, we have prepared new dose estimates that should be considered trustworthy in part because the greatest departure from previous dose estimates has been explained in part because of the use of modern techniques, and in part because the neutron levels have been shown to agree with *in situ* activation measurements. We have applied the Rossi and Mays leukemia methodology to these data and find that in this specific example the results no longer support an inference of large neutron RBEs. This is a dramatic example of the impact of our new dose values on estimates of radiobiological effects.

For further information contact William E. Loewe (415)422-6504.

Notes and References

1. W. E. Loewe and E. Mendelsohn, "Revised Dose Estimates at Hiroshima and Nagasaki," *Health Physics* (in press), also available as Lawrence Livermore National Laboratory Preprint UCRL-85411 (1980).
2. D. A. Auxier, *Ichiban Radiation Dosimetry for the Survivors of the Bombings of Hiroshima and Nagasaki*, Oak Ridge National Laboratory Rept. HP-27080 (1977).
3. H. H. Rossi and C. W. Mays, "Leukemia Risk from Neutrons," *Health Physics* 34, 353 (1978).
4. E. Straume and R. E. Dobson, "Implications of New Hiroshima and Nagasaki Dose Estimates: Cancer Risks and Neutron RBE," *Health Physics* (in press).



In many of the Laboratory's programs—the nuclear explosives test program, the laser fusion program, and the hydrodiagnostics testing effort for nuclear explosives design—the need to accurately characterize physical phenomena places continuing demands on our ability to make rapid measurements. In many cases, these measurements require that we resolve signals in the subnanosecond (10^{-9} s) range. These measurements involve both one-dimensional and two-dimensional (area imaging) recording of various physical phenomena and consequently require very high speed integrated circuits for associated signal processing and sensing.

Integrated circuits are built up of large numbers of transistors. Although a transistor may have many other applications, one of its most basic functions is as a simple switch, turning currents off and on. How fast it can do this (the rate at which it can respond to a signal) is a critical determinant of the ultimate value of the circuit in our high-speed applications. The fastest presently available silicon integrated circuits have switching times of about 500 ps (0.5 ns), enabling them to operate at frequencies of about 1 GHz.

A switching time of 500 ps is adequate for most applications but is much too slow for others. In fusion experiments for example, 500 ps can be a long time. Some diagnostic instruments require switches 50 times faster, operating in about 10 ps.[1] Electron streaking vacuum tubes do have very fast response times, but they use film as the recording medium, are limited to optical inputs, and are very expensive. New solid-state device developments will eliminate the need for film and also allow for the cost-effective integration of sensors and system electronics. Improving the

For further information contact Stefan P. Swierkowski (415)422-8206 or Lawrence E. Jelsma (415)423-0570.

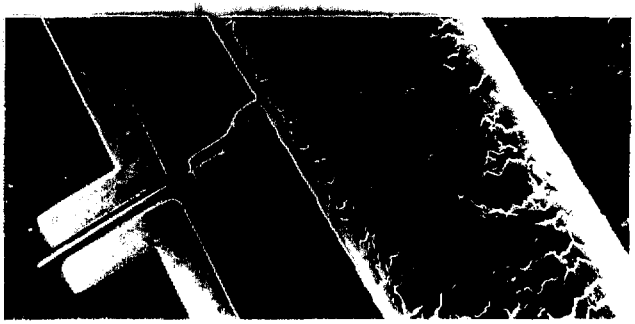


Fig. 1

A scanning-electron-microscope image of one of our prototype gallium arsenide MESFETs. The human hair to the right of the MESFET is 100 μm in diameter. The gate electrode (the narrow bar between the L-shaped source and drain electrodes) is 1.4 μm wide. The electrons flow from source to drain through the semiconductor beneath the gate electrode.

response time increases the instrument's resolution, which frequently reveals new physical phenomena, in somewhat the same way that the invention of the microscope revealed the subvisual world.

The major factor influencing switching time in a circuit is the speed at which charge carriers (and, hence, signals) travel in the semiconductor wafer. The primary way to improve switching speed is to minimize travel distances in the semiconductor.

However, such size reduction is already approaching its practical limits in the conventional contact photolithography used in fabricating integrated circuits. Circuit networks laid down by this method have conductors only 1 μm wide and 1 μm apart (a human hair is about 100 μm wide). Significant further speed improvements are prevented by silicon's properties, as discussed below. The next step is to find a better semiconductor than silicon.

Gallium arsenide (GaAs) has long been used for specialized transistors and Gunn diodes. Charge carriers move through it 5.5 times as fast as through silicon at relatively low electric fields, and it has other desirable characteristics. Small-scale integrated circuits made of GaAs have recently become available.

However, these commercial GaAs integrated circuits, designed to satisfy mass markets, fail to match our specialized needs. We are, therefore, beginning to produce our own GaAs integrated circuits based on the metal-semiconductor-field-effect transistor (MESFET, Fig. 1). We have developed a computer model that guides our development efforts by simulating the MESFET's behavior. The model, which can also predict circuit performance, indicates that MESFETs will have very short switching times.

Electron Transport in GaAs

To understand the special properties of GaAs MESFETs, we must know something about electron transport in semiconductors. The key parameters are the semiconductor's electron mobility, excess-carrier decay time, and bandgap.

In a simplified view, all the outer electrons surrounding the nucleus of an isolated atom are trapped in valence states (i.e., they have nowhere else to go). If the atom is incorporated into a crystal, the overlapping of the attractive potentials of adjacent atoms creates additional states (conduction states) in which electrons can move in response to electric fields. In describing mass motions of electrons, we can group the many electronic states into bands—a valence band and one or more *conduction bands*.

In a metal, the overlap of the conduction and valence bands creates a plentiful supply of electrons for current flow. In an insulator, there is a large energy gap (the bandgap) between the valence and conduction bands, no electrons cross over, and no current flows. In a semiconductor, the bandgap is narrow, various common processes can boost electrons across it, and applied voltages can affect the motion of the electrons. This ability to change resistance in response to a signal is what makes semiconductors so useful.[2]

In most semiconductors, there are multiple conduction bands. The principal conduction band (the one containing the largest number of mobile electrons) is the one that is lowest in energy. The conduction band may be thought of as a surface in electron energy-momentum space, and the electron mobility is strongly related to the curvature of the band surface at the state occupied by an electron.

So far, we have been describing electron transport in pure crystals. It is possible also to make a semiconductor material more conductive by doping it with selected impurities that contribute charge carriers.

One advantage of gallium arsenide over silicon as a semiconductor is that it

has a larger bandgap (1.43 eV compared to 1.12 eV). We can increase its conductivity in selected spots, and the rest of the crystal remains almost insulating. This greatly reduces parasitic currents that do nothing but generate unwanted heat and slow the circuit.

In addition to being narrower, silicon's bandgap is indirect. A conduction electron cannot decay, that is, drop directly back to the valence band; to conserve momentum it must simultaneously interact with a crystal lattice phonon (a quantum unit of lattice vibration). Gallium arsenide's bandgap is direct, enabling electron decay without a phonon. Disposing of excess electrons, as at the end of a pulse, is therefore much faster in gallium arsenide than in silicon.

In the case of a low electric field of 1 kV/cm (obtainable at 0.1 V across a gap only 1 μm wide), the electron velocity in high-purity silicon is 15.5 km/s. In high-purity gallium arsenide, it is 85 km/s.

In the case of high electric fields (above 3.5 kV/cm), the differences between silicon and gallium arsenide (illustrated in Fig. 2) are even more striking. In silicon's single conduction band, the electrons become "hot" (depart from thermal equilibrium with the crystal lattice) and begin to excite phonons, dissipating some of their energy. Instead of going faster in proportion to the increase in field strength, they tend to approach a limiting velocity.

Gallium arsenide has two conduction bands separated by an energy gap of 0.38 eV. Electrons drift much faster in the lower conduction band than in the upper band. As hot electrons form in the lower band, some of them scatter into the upper band, where they drift relatively slowly.

An energy gap of 0.38 eV is almost 15 times as large as the thermal (300 K) excitation energy of 0.0259 eV. Hence, few electrons move into the upper conduction band by accident. It takes energy to put them there.

For electric fields above abc 3.5 kV/cm, however, enough electrons have reached the upper conduction band to begin to reduce the average velocity. Thus, while the current at first increases with the field, it is possible in some devices for increasing fields to cause

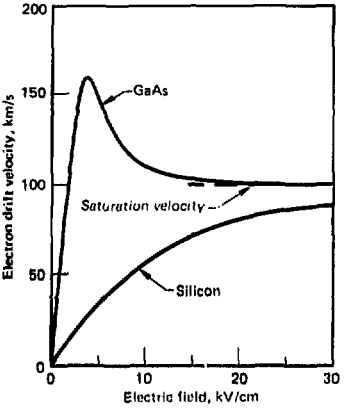


Fig. 2

The average drift velocity of electrons as a function of the electric field for gallium arsenide and silicon. The mobility is the ratio of drift velocity to the corresponding electric field. The drift velocities in both materials are similar at high fields but much higher at low fields in gallium arsenide than in silicon. Furthermore, the negative slope (where the velocity decreases for increasing electric field) in gallium arsenide above 3.5 kV/cm produces many useful nonlinear device phenomena.

decreasing currents. Operation in this region of higher fields is the basis of many useful phenomena such as the generation of microwaves in the Gunn diode oscillator.

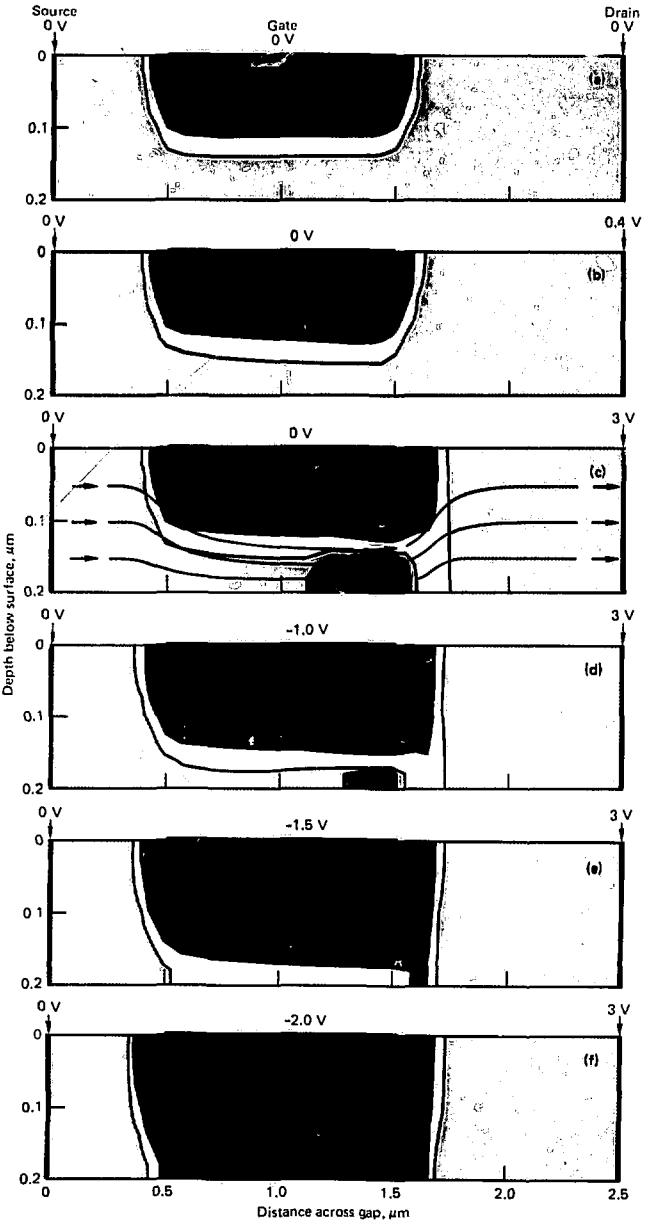
MESFET Construction and Operation

A MESFET consists essentially of a small area on the surface of a semiconductor crystal that has been selectively doped with impurity atoms and provided with three parallel conductors. See Fig. 1 and Fig. 3 for details of MESFET construction and operation. Either of the two wide outer conductors can be a low-resistance source of electrons (the second conductor acts as the drain), and the narrow electrode between them serves as the gate (or control) electrode. At low applied voltages, the current is directly proportional to the voltage difference between the drain and the source. Both the source and the drain are in good ohmic contact with the crystal, but the gate electrode makes rectifying contact, i.e., current would flow through it only if we were to give it a positive potential. Furthermore, this rectifying contact provides a chemically built-in bias voltage (the Schottky barrier effect) of −0.95 V.

The MESFET can be operated as a variable load resistor, a useful application for improving the frequency response of

Fig. 3

Simplified cross sectional diagrams illustrating MESFET operation as a nonlinear resistor (a to c) and as a fast switch (d to f). Contour lines indicate the loci of points at which the electron concentration is 115, 105, 100, 90, 50, and 5% of its equilibrium doping value. In (a), the built-in negative potential on the gate electrode repulses the drift electrons, depleting the colored areas and exposing the immobile positive ions (implanted donors). Increasing the drain potential increases the current flow through the bottleneck, depleting more of the region (b). At velocity and current saturation, the combined drift and diffusion current (1.3 V cm) is independent of drain voltage (c). Applying increasingly negative gate voltages cuts off the current completely as strong depletion chokes off the channel (d through f).



an amplifier. In this application, the source and gate electrodes are connected, giving them the same applied potential (zero). Nevertheless, the built-in negative potential on the gate may be thought of as repulsing the electrons that are flowing in the channel, confining them to a thinner layer and consequently increasing the local resistance (Fig. 3a).

As the drain voltage increases, the resistance under the gate builds up (Figs. 3b–c). More and more electrons are forced to flow through the restricted channel under the gate, which becomes choked with traffic to the extent that the electron velocity eventually levels off. In this condition, the drain current is almost independent of drain voltage.

The channel, in this saturated condition, may appear to be pinched off. However, the material is only partly depleted of electrons. As indicated by the arrows in Fig. 3c, large drift and diffusion currents of electrons can still flow.

If we now disconnect the gate from the source and change the gate voltage (Figs. 3d–f), the MESFET behaves like a switch. Even though we keep the source and drain voltages constant, the current to the drain can be reduced essentially to zero, i.e., switched off, by applying a sufficiently strong negative voltage to the gate (with reference to the source). In this case, the repulsive gate field pushes all the electrons away, effectively converting the channel into an insulator.

The MESFET is an extremely versatile device, serving as much more than a simple switch or variable resistor. Depending on how MESFETs are interconnected with each other, they can also function as amplifiers, modulators, or oscillators. Very complex circuits can be built up from arrays of MESFETs, all looking much alike but serving many different purposes.

The above simplified decription of MESFET construction gives little notion of the intricate procedures involved in translating a basically simple concept into a working device. Special washes, rinses, etching processes, ion implantation, annealing, and, above all, extreme cleanliness contribute to a successful MESFET. The slightest change in operating procedure or chemicals can

Charge carriers move through gallium arsenide 5.5 times as fast as through silicon at low electric fields.

increase the rejection rate from a few percent to 50% or more.

One reason this process is so critical is that the semiconductor's fixed surface charge density (the density of charges permanently attached to the free surface between the electrodes) is a function of the chemical and mechanical history of the semiconductor's surface. The attractive or repulsive forces of fixed surface charges can be equivalent to external gate potentials of several volts. Hence, small changes in chemical processing can produce large changes in operating characteristics.

Such difficulties are troublesome enough in a production plant devoted to making millions of identical devices. They are greatly compounded in a research application producing small batches of experimental designs. We must thoroughly understand the underlying principles if we are to predict the outcome of any changes we introduce from one version to another. An indispensable learning tool in achieving this understanding is a practical computational model.

The alternating difference implicit (ADI) scheme . . . is attractive because it is efficient in terms of computer storage and because it exhibits a high degree of numerical stability.

The Computational Model

To model MESFET behavior accurately, we must consider additional factors and transient effects. Carrier diffusion is proportional to the gradient of carrier concentration, and both the diffusion coefficient and the mobility are dependent upon the electric field. We may also have to include the finite lifetimes of carriers and (if we wish to model photoelectric response) the transport of holes.

In our computer model, we programmed Poisson's equation and the conservation equations for electrons and holes in difference form. This set of equations implies a quasi-static approximation, which is equivalent to assuming that time-varying magnetic fields from the MESFET's currents induce negligible electric fields. Because the MESFET is physically much smaller than the wavelength of the highest operating frequency, this assumption holds true for typical device operating parameters.

In our computer program MESADI (metal-semiconductor alternating

The Alternating Difference Implicit (ADI) Method

In general, our mathematical modeling is a two-stage process. In the first stage, we start from a convenient (but unrealistic) assumed initial-condition solution and calculate a realistic solution that meets the initial boundary conditions (potentials on the various electrodes, charge carrier concentrations, etc.). In the second stage, we change the boundary conditions and calculate how that affects the solution.

We can stop the calculation at the end of the first stage if we want to calculate only the steady state of the system. The first stage uses a pseudotime, and its transient solutions are essentially meaningless because they refer to the transition from an unrealistic initial condition. It is only in the second stage that time and the transient solutions are physically meaningful.

The alternating difference implicit (ADI) computational method further divides each stage into an alternating series of much simpler calculations. First, it separates the difference equations into two sets, one that pertains to the boundary conditions at the left and right and one that concerns top and bottom boundary conditions. Then, starting from an assumed (and possibly very poor) solution over the entire grid, the method solves one set of difference equations for the first row of nodes across the grid, to satisfy the left and right boundary conditions.

The computer approaches the solution of this problem stepwise, making closer and closer approximations, until it satisfies the boundary conditions within less than a certain limit (a millionth of the initial value, for example). When it has reached this limit, it applies the same procedure to the second row, and so on through the entire grid. Then, it does the same thing for each column of nodes, one at a time, using the second set of difference equations. It repeats this row-column iteration until the solution meets all the boundary conditions simultaneously.

At this point, the computer has a real steady-state solution that satisfies the initial boundary conditions. It then changes the boundary conditions (instantaneously changes the potential on some electrode, for instance), assumes a very brief lapse of time, and (using the same procedure but starting with the steady-state solution as the new initial condition) calculates a new solution that fits the new boundary conditions. When it reaches this solution, it assumes another brief time lapse and repeats the process again, eventually arriving at a new steady state. The intermediate solutions correspond to frames in a movie depicting the transient behavior.

The ADI method saves computer memory space and so produces solutions in much less time (and sometimes solves problems that no other method can). On a grid with 50 rows, 50 columns, and 3 variables, for example, alternative methods would require storage of 150 × 150 unknowns; with the ADI method, the corresponding storage is only 150 unknowns at a time. Furthermore, with the equations divided into two sets, we can formulate the problem in such a way that it can be solved more efficiently.

difference implicit), we have used the alternating difference implicit (ADI) scheme[3] (see box on this page). This method is attractive because it is efficient in terms of computer storage and because it exhibits a high degree of numerical stability. In this procedure, we replace the cross section of the MESFET with a computational grid and calculate the values of the unknowns (the potential and the electron and hole concentrations) at the grid nodes only. This strategem, however, would be impractical for a grid with a uniformly-spaced mesh; a constant mesh fine enough to give adequate detail in regions where the unknowns change

rapidly would have far too many nodes for the available computer storage. The solution is to use a grid with variable spacing in order to concentrate nodes in the places where fine detail is important.

Similar considerations apply to the time scale. When conditions are changing rapidly, we must calculate the changes over short time steps; as steady-state operation is approached, we use much larger time steps. However, it is difficult to calculate the point at which it might be desirable to change the time step. Therefore, we do the calculation in stages and decide at the end of each stage whether to increase the time step. In my case in calculating the changing potentials in a GaAs MESFET over an operating cycle of a few tens of picoseconds, we use many hundreds or even thousands of numerical time steps. Thus, for practical two-dimensional calculations of semiconductor models, we need the largest, fastest computers available.

For practical two-dimensional calculations of semiconductor models, we need the largest, fastest computers available.

Results

Using the MESADI computer program and some simplifying assumptions, we have produced a nine-minute animated movie[3] that exhibits some of the major features of typical GaAs MESFET behavior, including the variable-resistor mode of operation. The simplifying assumptions used in making this movie were: uniform impurity doping in the channel down to a certain depth, perfect gallium arsenide purity and zero conductance below that depth, instantaneous applied voltage change on the electrodes (square input pulse), and source and drain electrodes applied to ends of channel instead of top surface. These assumptions were made because postulating more realistic conditions would have made the calculation more expensive and the resulting solutions harder to interpret. Figure 4 consists of selected frames from the movie, showing what happens to the space-charge density (a–d) and the potential after the drain voltage is changed instantaneously from zero to 3 V (e–h).

In Fig. 4, we have inverted the potential scale to emphasize the relationship of electron potential energy to location in the MESFET cross section; it is highest at the gate, which is the negative electrode.

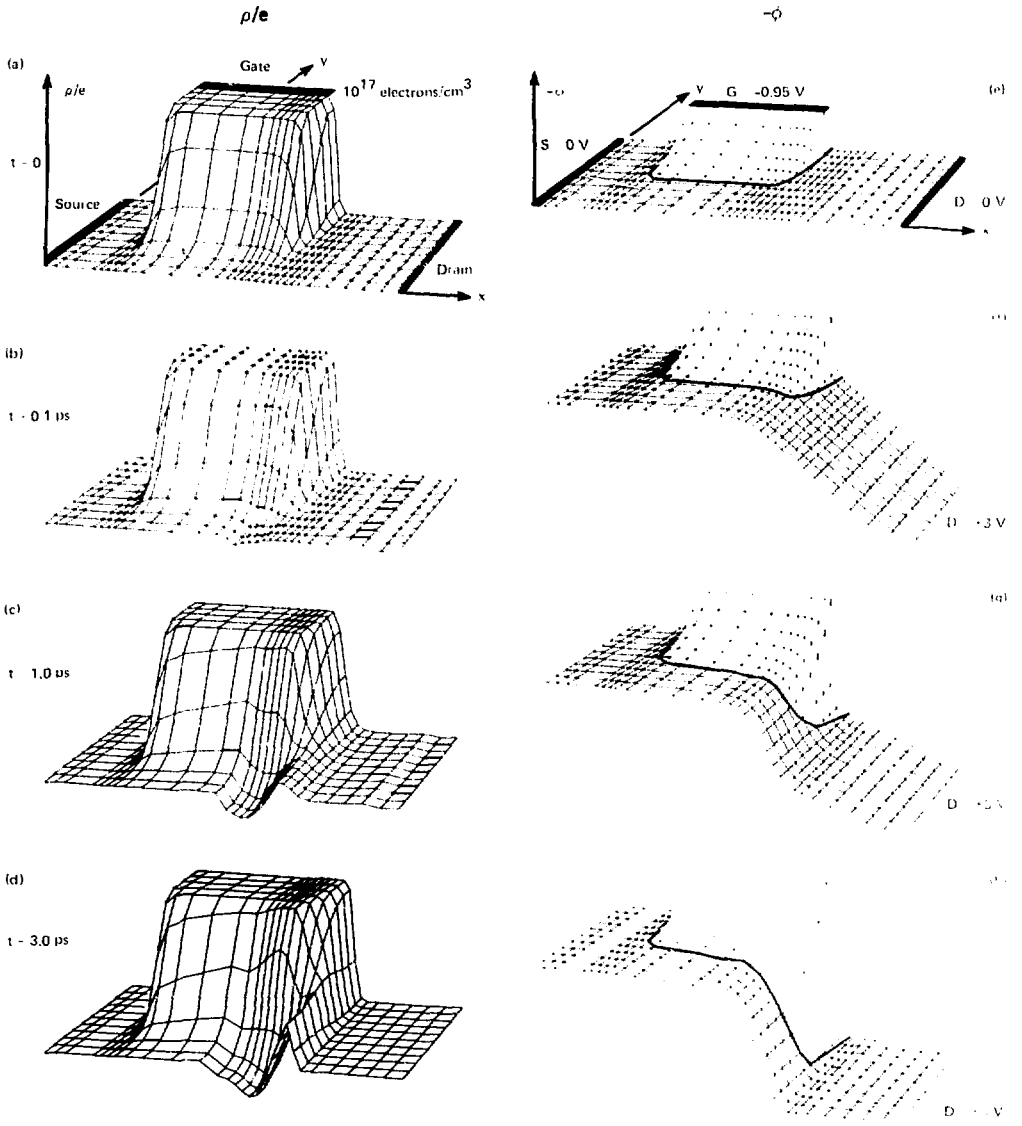


Fig. 4

Selected frames from an animated movie constructed by MISADI modeling to display the response of the space charge density (ρ/e plot, a through d), and the electrical potential (−ϕ plot, e through h) within the gallium arsenide MESFET (operated as a variable resistor) to a sudden change in drain potential. Most of the conduction takes place in the grey region of the plot shown in (e) through (h), where the electron concentration is highest. We have inverted the potential scale (negative potentials are plotted above the x-y plane) to emphasize the role of electron potential energy, which is high at the gate electrode.

The upper pair of frames shows the calculated initial steady-state conditions. Succeeding pairs show the development of the space-charge dipole (deep dip and rise at drain side of gate) and the potential distribution during the first 3 ps after application of the 3-V drain potential.

The thin depleted region directly under the gate is a potential barrier to electrons that enter the channel at the source and leave the channel at the drain. The concomitant space-charge density, normalized to the electron charge, summarizes the electron, hole, and donor-ion densities. The heavy line on the potential grid is the locus of points where the electron concentration is 50% of its value at thermal equilibrium. Carriers available for current conduction are confined largely to the green portion of the channel.

In the gate region, where the electrons are squeezed into a narrowed channel, the local field soon exceeds the 3.5 kV/cm threshold and rises into the strongly nonlinear portion of the velocity-field characteristic shown in Fig. 2. In this region, the higher electric field produces lower velocity, driving the field still higher. This factor is responsible for the extremely rapid build-up of the space-charge dipole. In this configuration, the potentials and electron distributions reach equilibrium in about 5 ps.

This 5-ps time corresponds to the charging time of an equivalent circuit using the distributed capacitance and resistance of the MESFET as determined independently by the model calculations, giving a straightforward cross check on the model results.

Comparing Calculations and Experiments

In the simplified calculation described above, the channel doping was assumed to be constant, and the semiconducting substrate was assumed to carry no current. A more realistic approach would model the ion-implanted donor distribution as a skewed Gaussian distribution—rather uniform near the surface and decaying rapidly with depth into the substrate. Such a distribution allows for some conduction current below the main channel.

Experimental measurements indicate that there is a fixed negative charge on the surface of the semiconductor between the electrodes. This charge will repel electrons even without help from an applied gate-electrode bias. It will also reduce the peak surface electric field that appears at the drain edge of the gate electrode.

When comparing the code results with experiments, we must allow for differences between the theoretical calculation and the experimental situation. In calculating the theoretical switching time, for example, we assumed an ideal input pulse (absolutely square with instantaneous rise time) and no stray capacitances. Neither of these conditions can be obtained in practice, where input pulses always have a finite rise time and stray capacitances are unavoidable. Furthermore, material preparation and device processing can affect the velocity-field characteristic of the material in the MESFET channel (see Fig. 2), so that calculations based on this characteristic are also uncertain.

Under such circumstances, it would be surprising if the theoretical and experimental values agreed within a factor of five. In fact, the theoretical response time for pinch-off is about 15 ps, and the best experimental value is 35 ps, although a more typical value is 50 ps. This agreement is quite reasonable, considering the experimental limitations.

The agreement between theory and experiment is even closer for calculations of steady-state results. By inserting a slightly reduced value of electron mobility derived from our experimental measurements, we calculated a drain saturation current within 10% of the experimental value.

The mechanisms underlying breakdown from excessive drain-to-gate voltages are poorly understood, but calculations with our model are beginning to indicate likely avenues of investigation. It has been conjectured, for example, that when breakdown occurs it is because of the high field associated with the space-charge dipole. However, our model calculations show that the local voltage gradients produced by this dipole effect are far too small, only about one-third of

the experimental bulk breakdown field of 300 kV/cm.

These same calculations, ignoring the effects of surface charge, locate the highest fields (about 600 kV/cm) at the drain edge of the gate. More complete calculations that included the effects of surface charge reduced this to 332 kV/cm, in reasonable agreement with the breakdown threshold.

Conclusions

The results of our MESADI code simulation indicate that several time constants are important in determining the transient response of a GaAs MESFET. The dielectric relaxation time, the charging time constant of the distributed capacitance, and the carrier transit time in the material each may play a significant role. Examination of the numerical solutions enables us to understand the relevance of each effect for a given device geometry and set of boundary conditions.

In the fabrication of MESFETs, we must control a large number of parameters such as line width, gap width, implant dose and profile, channel mobility (affected by processing temperatures), contact resistance, and breakdown voltage. Even the simplest integrated circuit involves at least five sequential lithographic steps. Surface chemistry, which can be affected drastically by minute changes in minor impurities in solvents or etch solutions, plays a major part in each of these steps.

Only by concentrating on the most critical processing parameters can we fabricate and test integrated GaAs circuits efficiently. Applications of this model can help to determine which of the many processing variables are the most critical. For example, the code results indicate that the gate electrode may be widened 50%, to 1.5 μm, without reducing the MESFET's effectiveness for certain applications.

Our first modeling efforts neglected the effects of free surface charge and of hole transport. More recent applications that did include the effects of surface charge enabled us to locate the point at which electrical breakdown is most likely to take place. We can also use the model to determine such lumped MESFET parameters as drain saturation current, pinch-off voltage, interelectrode capacitances, and the transconductance (change in drain current divided by change in gate voltage), enabling us to carry out circuit design calculations.

High-speed GaAs MESFET technology is characterized by microminiature gate widths and interelectrode gaps, high electric fields and mobilities, and short response times (less than 50 ps). The small dimensions prevent our measuring electric fields and current densities in enough detail for a thorough understanding of the observed electrical behavior. A two-dimensional numerical simulation makes it possible to envision what we cannot measure.

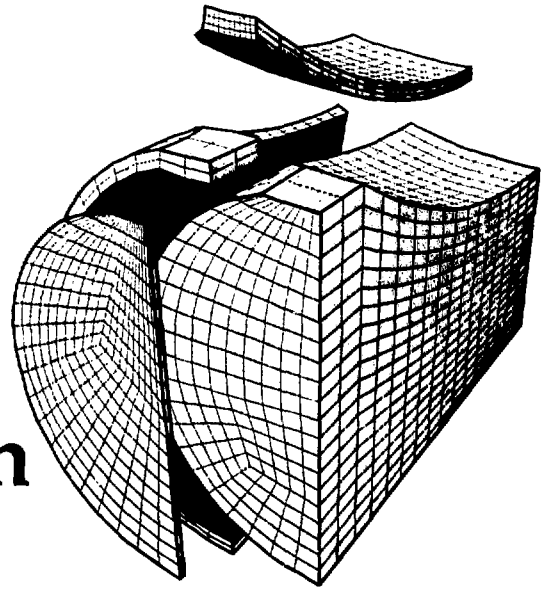
The GaAs MESFET technology is a leading contender for producing the high-frequency analog and digital circuits needed for state-of-the-art instrumentation. This technology will supplement rather than replace the very mature and flexible silicon device and integrated circuit technology. Its primary use will be for small-scale integrated circuits in applications where high-speed performance is of paramount importance. Two-dimensional numerical modeling is an important aid in our effort to develop GaAs MESFET technology.

Key Words: computer modeling, computer program, MESADI, MESFET, gallium arsenide, integrated circuit technology, semiconducting materials

Notes and References

1. For a discussion of the use of high-speed digital equipment to record electrical transients, see *Energy and Technology Review* for June 1979 (UCRL-52000-79-6), p. 5.
2. For background on semiconductor band structure, see *Energy and Technology Review* for April 1981 (UCRL-52000-81-4), p. 11.
3. A. R. Mitchell, *Computational Methods in Partial Differential Equations* (J. Wiley & Son, 1977), p. 103.
4. The film depicting MESFET behavior may be borrowed from the LLNL library as UCRL-52000-81-8-1.

DYNA3D: Modeling Material Deformation



DYNA3D, a finite-deformation, finite-element computer program, enables us to study in three dimensions material interaction and deformation due to applied forces.

For further information contact John O. Hallquist (4151422-8756).

DYNA3D is a computer program that enables us to study in three dimensions material interaction and deformation due to applied forces. We developed this program for several applications, none of which could be handled by existing codes. These applications require that we be able to:

- Study the structural integrity of LLNL-designed nuclear weapons after impact upon the ground but before detonation.
- Predict the structural failure of reentry vehicles due to impact, as part of a nonnuclear-kill program that is concerned with the destruction of reentry vehicles using high-velocity fragments.
- Design a class of shaped explosive charges, called self-forging fragments, with nonaxisymmetric shapes.
- Make three-dimensional studies of the interaction of nuclear reactor structures with the surrounding soil, as part of a seismic safety program.

We began to develop DYNA3D in 1976 and since then have generated many versions, each more efficient and more versatile than the last.[1] DYNA3D is a Lagrangian, finite-deformation, finite-element computer program that solves the momentum equations and the conservation equations of mass and energy. The applicability of this computer program is directly due to the finite-element method of performing complex computations.

The Finite-Element Method

By using the finite-element method, we can divide a complicated spatial problem

into manageable parts. First, we organize the problem into an assemblage of discrete, box-shaped regions called elements. The location of each element is specified with respect to a coordinate system. This assemblage of elements is called a mesh. The behavior of each element is governed by a set of mathematical equations. Elements are interconnected at points called nodes, and each node is associated with three displacements (the x, y, and z directions). The mass of each element is distributed equally to its corner nodes.

Displacements are assumed to vary linearly over each element. This assumption allows us to find the displacement of any point within an element by interpolating the known displacements at the nodes. Displacement compatibility is maintained between adjacent elements. Thus, we can compute the material strain (i.e., deformation) at any point in an element, and from this we can determine the stresses acting on it and other elements.

Given the element stresses, we can calculate concentrated forces that act at the nodes. We obtain accelerations by dividing these nodal forces by the nodal masses. Then, we perform explicit time integration to update velocities and displacements. In this way, we can learn how the overall structure behaves.

DYNA3D uses a Lagrangian formulation that allows the mesh to conform to the material boundaries. As the materials move, the mesh also moves, and the size and shape of the elements adjust to fit the changes in the materials. The box on p. 15 describes the finite-element method in more detail.

Features of DYNA3D

The material models implemented in DYNA3D enable us to study many different materials. These models include a model of elasticity; plasticity models that allow thermal softening, pressure hardening, and work hardening; fracture, soil, and foam models that have multiple yield surfaces and volumetric crushing; and models of reactive and nonreactive high-explosive equations of state. In addition, DYNA3D can easily accept new material models of great complexity.

To illustrate the application of DYNA3D's material modeling, in Fig. 1 we show a computer analysis of a nuclear reactor structure embedded in a very nonlinear soil and subjected to an earthquake. Because of the complexity of the soil's behavior, this model uses 78 variables per mesh element (instead of the eight or nine used in simple, equivalent linear models). The purpose of this study was to determine whether very detailed and complex models of soils predict significantly different structural responses from those obtained with simpler soil models.[2] We found that, while simpler, equivalent linear models can properly simulate low-frequency stress transmissions, complex nonlinear models are needed to accurately model higher frequency stress transmissions.

Another important feature of DYNA3D is that it can model the behavior of two materials sliding past each other. DYNA3D uses special-purpose algorithms to model the abrupt

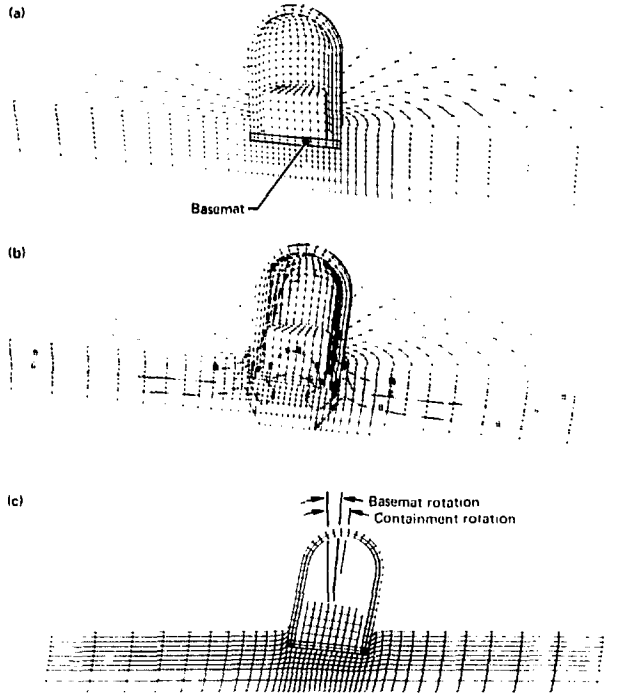


Fig. 1

Earthquake analysis of a nuclear reactor using DYNA3D and a very detailed soil model. This was one of a series of analyses done to determine whether the structural response of the reactor building changes significantly as the soil model becomes increasingly sophisticated. (a) The finite-element mesh of the starting condition; (b) contours of shear stress; (c) the positions of the basemat and the reactor containment structure after the earthquake.

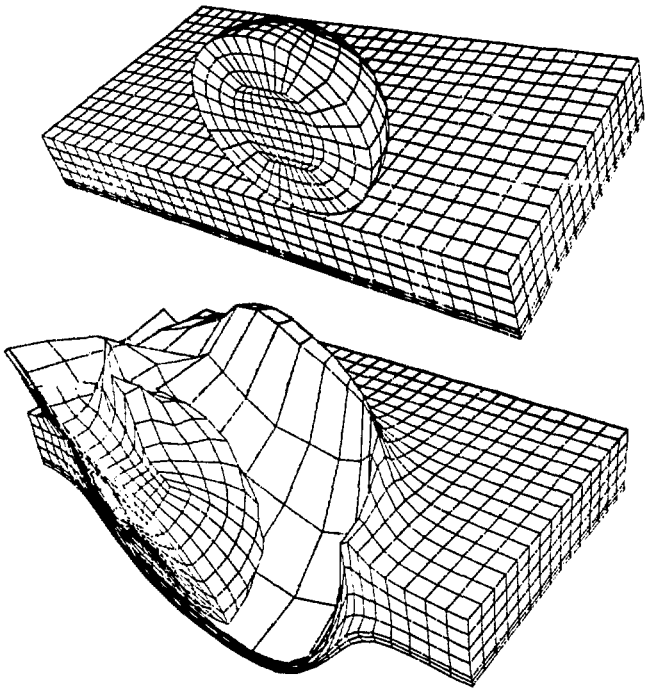


Fig. 2

DYNA3D analysis of the impact of a solid aluminum ball on a mock reentry-vehicle casing (an aluminum plate covered with a four-element-thick layer of carbon phenolic heat shield) at an angle of 45 deg and a velocity of 6 km/s. A sliding interface between the ball and the casing permits the calculations to be properly computed. The figure shows the deformed shape 3 μs after impact.

changes in velocity and other physical conditions that can occur at an interface between materials. Without this capability, the usefulness of DYNA3D would be limited because when all parts of a structure are modeled as if they were tied together (as is the case in most finite-element codes), the structure appears to be much stronger than it really is. With these algorithms, DYNA3D also can avoid creating such phantom results as tensile stress between materials separating at an interface and interpenetration of parts under compression.

Figure 2 shows the DYNA3D modeling of an aluminum ball impacting obliquely on an aluminum plate that is covered by a carbon phenolic heat shield four elements thick. Without the algorithms for sliding material interfaces, the calculation would be pointless because the ball would not "know" about the existence of either the heat shield or the aluminum plate. Calculations like the one from which Fig. 2 was derived are run routinely to check for anomalies in the sliding interface algorithms.

DYNA3D can model a variety of loading conditions. Loads may be approximated in the calculation by applying pressures on free surfaces, by specifying nodal-velocity time histories of a subset of boundary nodes, by applying body-force loads due to base accelerations and angular velocities, and by specifying initial velocities. Loads can be generated also by linking DYNA3D to another computer program, the three-dimensional Eulerian hydrodynamics code, JOY.[3] In addition, loads can be generated by calculating high-explosive detonations. With DYNA3D, there is almost no limit on the number or type of loading conditions that may be used in a given problem.

Because DYNA3D cannot generate meshes, it relies on other computer programs to generate the finite-element meshes. For example, many problems with nonaxisymmetric loads can be calculated on an axisymmetric mesh, which can be created with a two-dimensional mesh generator and then "spun" around the axis of symmetry with the SPIN program.[4] SPIN also automatically generates all the definitions for the sliding material interfaces. Meshes that are nonaxisymmetric can be generated with the SLIC program.[5]

To permit complete analysis of the calculational results (post-processing), DYNA3D creates two data bases, one to record the frequently taken time-history data for the nodes and elements of special interest (specified at the beginning of the run) and the other to record the less frequently taken node and element data for the entire mesh. Post-processing of the time-history data can be done with DYNAP.[1] Geometry plots, with or without contours, can be generated with GRAPE,[6] which can also be used to generate continuous-tone color pictures and to contour any given variable in color.

The data-storage capacity of DYNA3D is very large. Currently, the program can store 80 000 elements on the

Cray-1 computer if all element data are on disk and 30 000 elements if the element data are stored in core. Because transferring data between disk and core is relatively costly in computer time, we do most of our DYNA3D calculations at LLNL with the element data stored in core. However, for facilities with computers that have much less memory than the Cray-1, DYNA3D's capability to store data on disk permits much larger problems to be run than would otherwise be possible. On machines that are slower than the Cray-1, it is possible to overlap DYNA3D calculations with data transfers to and from disk; thus, the speed of the program is not severely decreased by storing element data on disk.

DYNA3D has been fully vectorized for the Cray-1 computer. That is, vectors (long arrays of numbers) rather than scalars (single numbers) are used to perform all important computational operations. DYNA3D can run nearly four times faster on the Cray-1 using vectors than it would using scalars. However, vectorizing a computer program can be difficult since many operations cannot be vectorized. To avoid this problem, we have recently reformulated many algorithms in DYNA3D to eliminate these nonvectorizable operations. By so doing, we have attained very fast execution speeds on the Cray-1, more than a hundred times faster than the 1976 version of DYNA3D on the CDC 7600. Currently, it takes only 40 μs to compute one element over one time step, making DYNA3D the fastest program of its type in existence.

Using DYNA3D to Analyze the Structural Design of the B83

The B83 is a complex weapon that must be able to function after surviving a severe delivery environment. The bomb is designed for laydown deployment, that is, to be released from a low-flying aircraft, and, after parachute retardation, to impact the ground at velocities of about 33 m/s. The bomb must survive impact on hard, irregular objects. The detonation is delayed to permit the delivery aircraft

The Finite-Element Method

Before a computer program can calculate a structure's response to applied forces, it must be provided with a mathematical description of the object. Simple mathematical formulas usually are not satisfactory for this purpose because they can fit only regular geometrical shapes and so can allow no significant distortions. Instead, we must divide the structure being represented into a large number of elements and specify where each element is with respect to a coordinate system. This process is called discretizing. Then, the computer program can calculate how the applied forces affect each element and how each element affects its adjacent neighbors. A master calculation summarizing all these individual calculations can describe what happens to the whole structure.

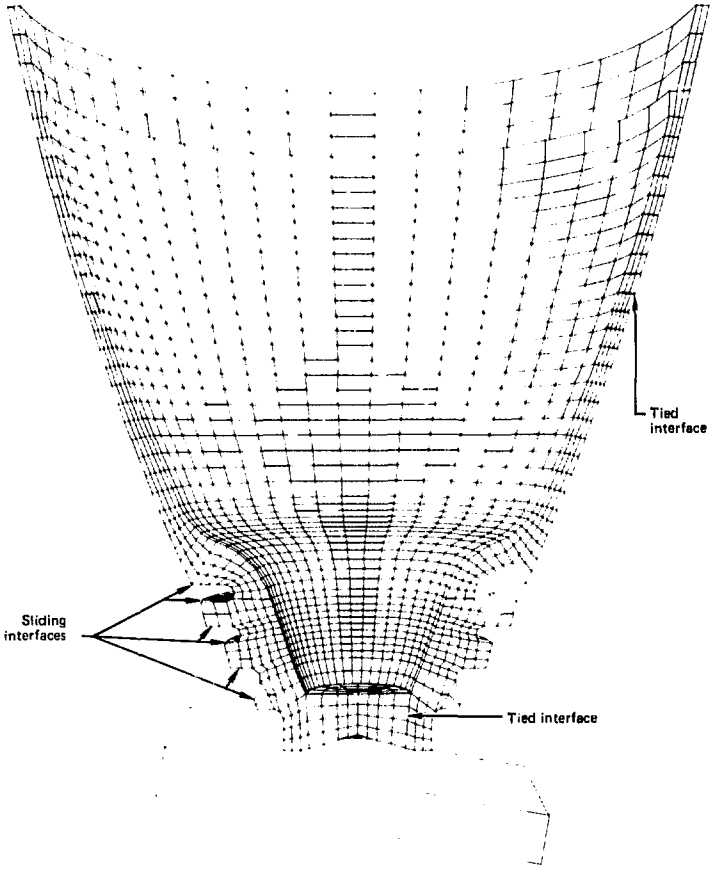
Elements are shaped somewhat like bricks, and their shapes are defined by the positions of their corners (called nodes). Elements must be quite small to map out the entire structure without seriously distorting either its geometry or the temperature, pressure, and other gradients inside it. However, the smaller the elements are, the more of them there will be and the more complicated, time-consuming, and expensive the calculation will be.

To obtain a complete picture, the calculation must follow a structure's response to the applied forces through time. Thus, the time-step factor is an important consideration in generating a finite-element mesh. The program calculates how each element affects its neighbors, but only its adjacent neighbors. Hence, the program must be limited to calculating in time intervals so short that no disturbance in any one element can propagate beyond its adjacent neighbors.

This sounds simple, but in practice it is quite complicated. A disturbance travels at the speed of sound, and this speed varies as a function of the temperature and density of the material. Even if the elements are all the same size at the start, once the structure begins to deform, some elements may be squashed flat or stretched thin. Sound can cross a thin element much more quickly than a thick one. Therefore, the program must search through all the elements, find the one in which the sound transit time is the shortest, and make the time step for the next calculation less than this minimum transit time. Obviously, it is possible for the time-step size to become very small, with the result that the calculation will require many time steps and a great deal of computer time.

Fig. 3

Finite-element mesh for the DYNA3D analysis of the B83 stainless steel crushable nose. The top row of elements simulates the aft end of the bomb, and an initial velocity of 5.75 m/s is assigned to each node. Multiple sliding interfaces and two tied interfaces permit transitions in the mesh (see text, last paragraph on p. 15). The upper tied interface reduces the number of elements through the shell thickness from 4 to 3, and the lower tied interface reduces the number of circumferential elements from 20 to 8 (enabling us to simplify the calculation and save much computer time). The bottom element models the rigid wall upon which the bomb impacts at a 10-deg angle. This was a complex calculation requiring 4356 elements and 6074 nodes.



to escape. Much of the early incentive to develop DYNA3D arose from the need to understand the broad range of environments that might be encountered when a bomb is deployed in this way. The wide variety of materials in the B83 bomb, including steel, high explosives, and crushable steel-honeycomb shock mitigator, was beyond the modeling capabilities of existing programs.

We are working with Sandia National Laboratories, Livermore, to apply DYNA3D in structural design calculations of nose-on impact conditions for the B83. The crushable nose is being redesigned to meet new and more demanding requirements for absorbing kinetic energy upon impact. To validate the application of DYNA3D to the bomb design process, we have compared computer calculations with the results of a Sandia experiment on a one-fifth scale model.[7] The use of DYNA3D in the nose-crush analysis offers the potential of a considerable reduction in design time, especially since it takes nearly four months to fabricate and test a redesigned part.

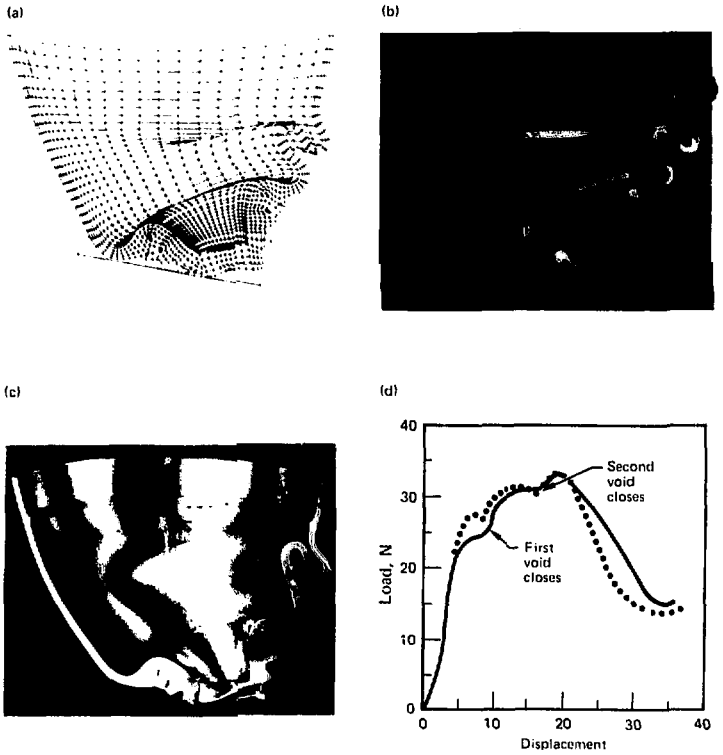


Fig. 4

(a) The DYNA3D nose-crush calculation and (b) the contoured model showing regions of plastic strain (color contouring done by the GRAPE code shows blue, cyan, green, yellow, and red corresponding to plastic strains of 0, 12.5, 25, 37.5, and 50%, respectively) compared with (c) the experimental nose crush. Agreement is excellent. In (d), the curves for load vs displacement for the computer analysis (solid line) and the experiment (dotted line) are compared, and, again, agreement is very good.

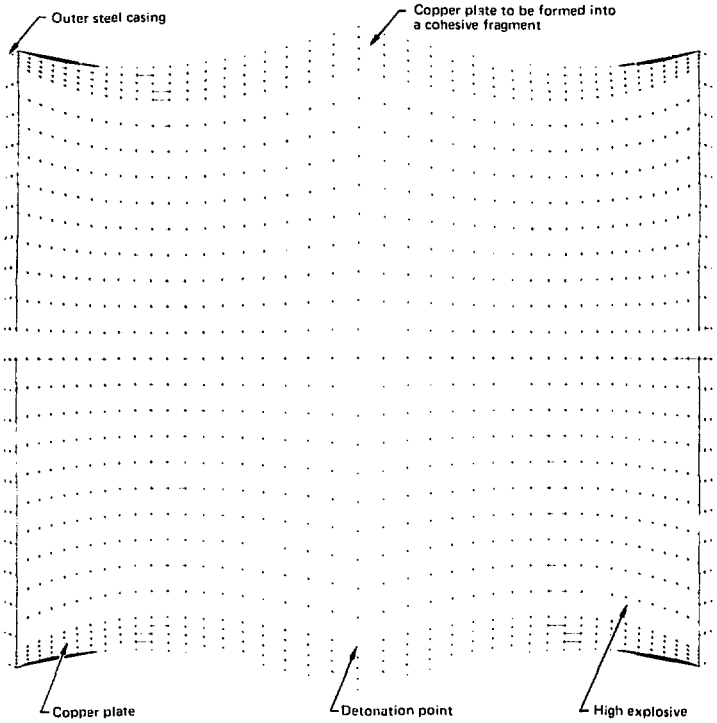
The finite-element mesh shown in Fig. 3 was used as the starting point in our analysis of the B83 nose. The mass of the bomb is simulated in the top row of elements; an initial velocity of 5.75 m/s is assigned to each node. The element at the bottom models the rigid wall. The bomb impacts this wall at a 10-deg angle. In the DYNA3D calculation, we model the 15-ms interval before the nose rebounds. Figure 4 shows the computed and experimentally deformed shapes of the B83 nose, together with the curves for computed load and measured load vs displacement. The agreement is strikingly good.

Designing Self-Forging Fragments

Self-forging fragments, also called Misznay-Schardin munitions, are being designed as antiarmor devices. This munition uses high explosive to accelerate a metal plate to a high velocity. The plate is designed to assume a cohesive shape during acceleration that enables it to penetrate thick armor. This acceleration-induced shape must also provide sufficient aerodynamic stability to carry the projectile over long distances. A plate with all the desired properties can be designed by iteratively modeling a variety of possible designs. A computer

Fig. 5

Three-dimensional finite-element mesh for the axisymmetric self-forging fragment calculation. A 90-deg segment is modeled with two symmetry planes. Sliding interfaces are placed between the high explosive, the outer casing, and the upper and lower plates. The mesh consists of 6% elements.



prediction of the behavior of an object will be only as accurate as the models of its constituent materials that are used. A plate made from the computer design is tested to ensure that it performs as predicted.

The two-dimensional program DYNA2D[8] has been used at LLNL in work on axisymmetric designs.[9] In order to assure ourselves that DYNA3D can do useful calculations, we repeated a two-dimensional DYNA2D calculation in three dimensions by spinning the DYNA2D mesh 90 deg with the SPIN code, which produced the mesh shown in Fig. 5. With this mesh we calculated the behavior of the plate until 50 μs after detonation of the high explosive, by which time the pressure of the explosive had dropped two orders of magnitude (from 30 GPa to less than 0.30 GPa).

Since the explosive would have only a negligible effect on the final shape of the fragment after 50 μs, we deleted all elements except those in the upper plate at this point. We then resumed the calculation to 200 μs to achieve a stable configuration for the fragment. The two sequences in Fig. 6 depict the results of both stages of the calculation. Figure 6(a) shows the detonation and expansion of the high-explosive gases to 50 μs, and Fig. 6(b) the deformation of the fragment to 200 μs. The considerable plastic strain, apparent in the second sequence, necessitates a sophisticated material model to calculate the extreme distortion that occurs.

The DYNA2D and DYNA3D computations agreed except for minor and easily explained differences. This calculation gave us confidence in our

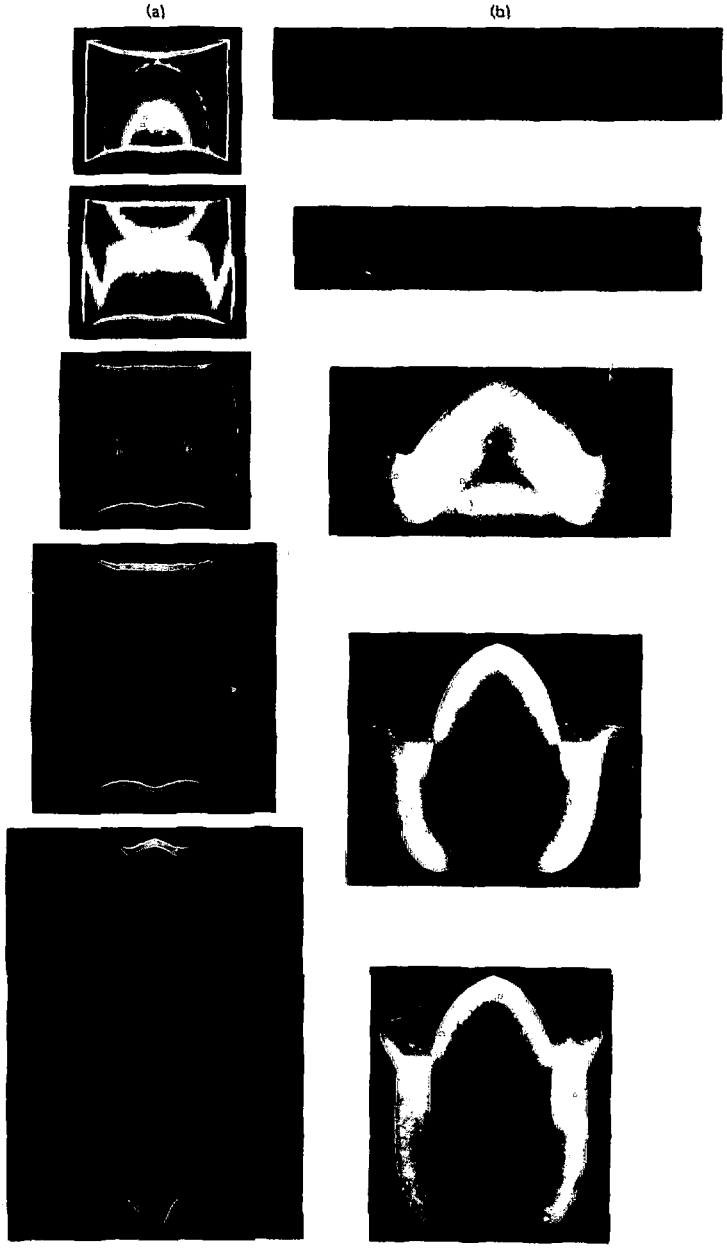
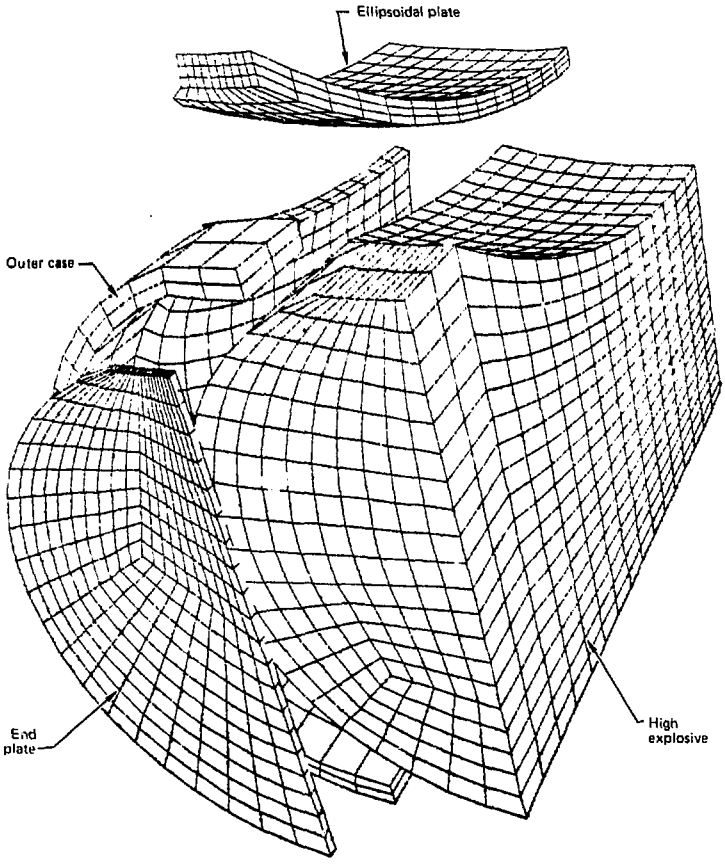


Fig. 6

(a) Detonation and expansion of the high-explosive gases at 5, 10, 15, 25, and 50 μs for the axisymmetric self-forging fragment. Regions in the high explosive where pressure exceeds 10 GPa are shaded red. Colors blue, cyan, green, and yellow correspond to pressures of 0, 2, 5.0, 7.5 GPa, respectively. (b) Deformed fragment at 0, 40, 80, 120, and 200 μs. Regions in the fragment where plastic strains exceed 200% are shaded red. Colors blue, cyan, green, and yellow correspond to plastic strains of 0, 50, 100, and 150%, respectively.

Fig. 7

An exploded view of a finite-element mesh of a cylindrical design of a self-forging fragment is shown with an ellipsoidal plate that ultimately will be turned into a fragment. One quarter of the design is modeled with two symmetry planes. Sliding interfaces are defined between the high explosive and the outer case, end plate, and fragment. Successful calculations have been completed with this mesh.



computations of fully three-dimensional geometries such as the one shown in Fig. 7.

Summary

DYNA3D was developed at LLNL to calculate material interaction and deformation under applied forces.[10] Calculations with this program have been compared with experimental data many times, with excellent agreement. DYNA3D is now being used with great confidence in the structural design of weapons systems, including the B83. We are also using DYNA3D in our nonnuclear-kill program to predict the structural response of reentry vehicles to impact loading, in our seismic safety program to analyze nuclear reactor structure , and in support of the Nonnuclear Ordnance Program's work on designing three-dimensional self-forging fragments. DYNA3D has been placed in the public domain and is operational at many computer installations throughout the U.S., Canada, and Europe. We are continuing our efforts to improve DYNA3D; in particular, we are planning to include additional equation-of-state models and to improve the material models.

Key Words: computer-aided design, computer analysis—three-dimensional, computer program (DYNA3D); finite-element method, material deformation, material interaction.

Notes and References

1. J. O. Hallquist, *Preliminary User's Manual for DYNA3D and DYNAP*, Lawrence Livermore National Laboratory, Rept. UCID-17268, rev. 1 (1979).
2. R. C. Chun *et al.*, *Uncertainty in Soil Structure Interaction Analysis of a Nuclear Power Plant. A Comparison of Linear and Nonlinear Analysis Methods*, Lawrence Livermore National Laboratory, Rept. UCRL-84190 (1981).
3. R. Couch is responsible for JOY, an Eulerian computer program currently being developed.
4. B. E. Brown, *SPIN—A Program for 3D Axisymmetric Geometry Generation*, Lawrence Livermore National Laboratory, Rept. UCID-18790 (1980).
5. M. A. Gerhard, *SIC—A Computer Graphics Mesh Generator for Finite Element and Finite-Difference Application Programs*, Lawrence Livermore National Laboratory, Rept. UCRL-52823 (1979).
6. B. E. Brown, *Displaying the Results of Three-Dimensional Analysis Using GRAPE, Part One: Section Graphics*, Lawrence Livermore National Laboratory, Rept. UCID-18507 (1979).
7. M. L. Chiesa and M. L. Callabresi, "Nonlinear Analysis of A Mitigating Steel Nose Cone," *Computer and Structures* **13**, 295 (1981).
8. J. O. Hallquist, *User's Manual for DYNA2D—An Explicit Two-Dimensional Hydrodynamic Finite-Element Code with Interactive Rezoning*, Lawrence Livermore National Laboratory, Rept. UCID-18756 (1980).
9. R. W. Werne *et al.*, *An Advanced Analytical and Experimental Investigation of the Hydrodynamic Performance of a Missile Schaudry Warhead*, Lawrence Livermore National Laboratory, Rept. UCRL-83406 (1980).
10. DYNA3D was developed by J. O. Hallquist. G. T. Goudreau calculated [illegible] of the nonaxisymmetric self-forging fragment used in this article.

Oil and Gas Reserves:

Predicting the Unpredictable

The 1973 oil embargo caused great concern about the size of this country's (and the world's) remaining fluid hydrocarbon resources, that is, the amount of recoverable oil and gas remaining to be discovered. There was considerable uncertainty about the actual amount, and official estimates differed by large factors. Within a few years, with considerable effort on the part of analysts, the estimates of what will be found in both old and new fields had begun to converge on a range of about 100 to 120 billion barrels for the U.S. as a whole (Table 1). Comparable estimates for the world are much more speculative and are on the order of 1000 billion barrels. How much exists is only part of our concern, however. The question of greater concern is, how much will be available on the open market.

Only a small part of the answer to this question is technical. Of much greater importance are the political and economic consequences of the actual location of these resources. Oil found under 1000 m of water and another 50 m of shifting pack ice may well be prohibitively expensive to produce. Governmental and political considerations may play an even larger role in determining the flow of oil and natural gas. Oil that is under the control of an unstable government or one that is at war may be essentially inaccessible. Oil needed for the producing country's own development probably will stay at home rather than find its way into the world's pipelines.

Table 1 Estimated U.S. reserves and resources of crude oil and natural gas liquids, billions of barrels.

Source[a]	Discovered		Undiscovered
	Proven or demonstrated reserves	Inferred or indicated additional reserves	Mean[b]
National Petroleum Council (1973) estimated as of 1-1-71	47	5	151[c]
Hubbert (1971) estimated as of 1-1-72	71		72
National Academy of Sciences (1975) estimated as of 1-1-75	31	5	113
Moody & Esser (1975) estimated as of 1-1-75	31[d]		50 (150 (85)[e]
Miller et al. (1975) estimated as of 1-1-75	35 30[f]	29[g] 23	e1 (140 (97)[f] 50 (127 (82)[f,h]
Langston (1976) estimated as of 1-1-75	41	50[i]	e8 (198 (118)[j]
Rand Corp. (1981) estimated as of 1-1-76	48 32[k]	17–72 1e–ee13e[l]	18–38 11–32 (120)[k]
USGS (1981) estimated as of 1-1-81			e4–105 (63)[m]

[a]See Reference 2 for source citations.
[b]Where a range is given, it is followed by the mean in parentheses.
[c]American Petroleum Institute estimates as of date indicated.
[d]Assuming 40% recovery of undiscovered oil-in-place (385 × 10⁹ bbl); water depth 450 m.
[e]Water depth to 200 m for reserves and 1800 m for undiscovered potential.
[f]Crude oil only.
[g]Water depths to 180 m.
[h]Includes offshore areas to 1800 m depth and marginal deposits currently uneconomic.

Oil-industry planning must be long range because of the long lead times involved in constructing massive facilities associated with off-shore and Arctic exploration and development. Long-range forecasts of oil and gas prices are also important in planning because it is these prices against which alternative energy sources will be competing. When oil supplies are governed primarily by market pressures and technical factors, long-range planning is feasible though difficult. When political and economic-policy factors enter the picture, long-range planning becomes exceedingly uncertain.[1]

Explorations and Discoveries

The uncertainties associated with estimating the amount of oil still to be discovered are small compared with the uncertainties involved in forecasting the political and economic forces that actually control final deliveries of petroleum. For example, consider the problem of predicting, even as much as six months in advance, any one of the major developments in the recent history of the countries surrounding the Persian Gulf. To illustrate some of the difficulties involved, we have examined some recent petroleum discoveries and attempted long-range predictions of how these will affect supplies.

Mexico

In 1921, Mexico produced 0.55 million barrels/day (mb/d) of crude oil. Production then declined and remained below 27 mb/d for 35 of the next 40 years. By 1971, production had gradually increased almost to the 1921 level. Suddenly, discoveries at Reforma, Chicontepec, and the Gulf of Campeche released a flood of oil. In the past 10 years, production has risen to almost 2.5 mb/d (Fig. 1). Only the U.S., the USSR, and Saudi Arabia currently produce more. Mexico's oil and gas reserves are comparable in extent to those of Kuwait, Iran, and the USSR, standing now at 67.8 billion barrels of crude oil and gas equivalent.

How does the presence of this large oil resource right next door affect the U.S. oil-supply picture? Less than one might suppose.

Currently about 65% of Mexico's crude oil exports come to the U.S., and Mexico's production is expected to nearly double (to 4.1 mb/d) by 1990. Her exports to us will not double, however. Mexico has announced a National Energy Program for the 1980s that limits annual exports to any single country to 0.75 mb/d (and total exports to twice that), only 2% more than we are receiving now.

This stringent export policy may or may not persist beyond the end of President Lopez Portillo's term in 1982, but another factor could also limit Mexican exports. Her annual domestic demand now stands at 1.3 mb/d and is growing by about 8% a year. At that rate, it would reach 3.4 mb/d by 1990, leaving about the same amount for export as there is today, despite a vastly increased production.

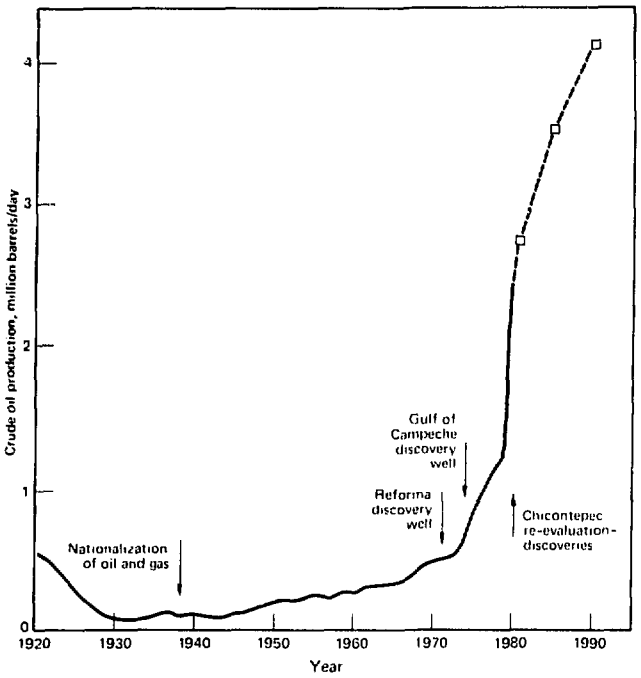


Fig. 1

Mexican oil production from 1920 to 1990. The great surge in production is due to discoveries in the Reforma area (in the Isthmus of Tehuantepec) and in the Gulf of Campeche. Projections (indicated by a dashed line and open squares) are from the Mexican National Energy Plan.

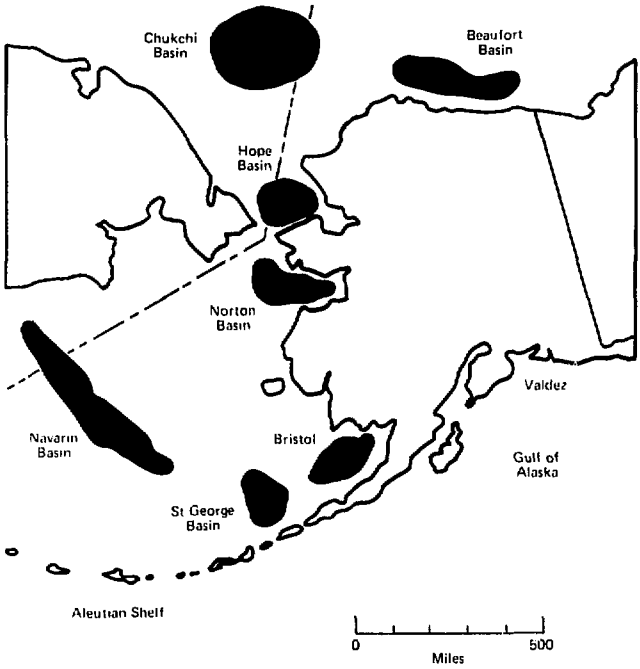


Fig. 2

Potentially hydrocarbon-bearing sedimentary basins on the continental shelves surrounding Alaska. Figures indicate U.S. Geologic Survey estimates of the reserves for each basin in billions of barrels of oil.

Arctic Ocean

Most experts agree that many outer continental shelf areas around Alaska (Fig. 2) are promising places to explore for oil because they contain sedimentary basins with hydrocarbon potential. Federal lease sales started in 1979, and others will follow, beginning in 1982. The Shell Oil Company has estimated that with accelerated leasing, production could be expected to approach 4 mb/d by 1995.

There are several reasons to doubt this optimistic forecast. Two of the richest of these basins, the Chukchi and the Navarin, straddle the border between the U.S. and the USSR, a circumstance that will complicate and could delay development. The short drilling season and the extremely hostile climate severely limit exploration activity here as elsewhere in the Arctic. Government observers question the capability of the oil industry to muster capital, manpower, and equipment in these virtually unpopulated areas in time to fulfill their production estimates for 1995.

Similar considerations apply to the Canadian Beaufort Sea area. There has already been at least one important discovery well and two years of exploratory drilling in the Beaufort-Mackenzie basin, which is estimated somewhat speculatively to contain 32 billion barrels of oil. The exploration and development of this basin will be extraordinarily expensive, however. One well, which is being drilled from an artificial island, will cost $75 million, almost four times as much as a typical well on the Alaskan North Slope.

Africa

A very large recent discovery off the Ivory Coast of Africa near Jacqueville consists of low-sulfur crude oil and has a gas cap to assist production. On this account, it has been compared to the Ekofisk field, whose discovery initiated the development of the North Sea oil and gas province.

To date, four holes have been drilled in this field, and exploration is expected to resume this summer. The water in this area is 360 to 450 m deep, and exploration is hampered by the lack of suitable deep-water drilling rigs.

Egypt

Egypt is a newcomer to the ranks of oil exporters (Fig. 3). Since 1975, when oil fields captured by the Israelis were returned to Egypt, production has increased by about 0.1 mb/d every year. Now at 1 mb/d, it nearly equals that of Kuwait, the bulk of it coming from newly developed fields in the Gulf of Suez.

This bounty is likely to be shortlived, however. Egypt's proven reserves total only 3 billion barrels, about eight times its annual production. The active exploration program is promising, but there is no way to estimate how successful it may be. Furthermore, supplies are jeopardized by the political instability of the entire Middle East.

Domestic Supplies

Can we forecast domestic production with more assurance than international production? Perhaps, but perhaps not.

The Overthrust Belts

The Western Overthrust Belt, now believed to extend from Alaska to Southern Mexico, appears to have been caused by a collision between two ancient continents. In the process, the two continents broke up and some pieces slid over others, producing multiple layers of hydrocarbon-bearing source rocks.

Present exploration was made possible by improved seismic methods that have identified drilling targets beneath what was previously believed to be thick basement rock. Recent U.S. Geological Survey estimates of recoverable oil in the nine fields in Wyoming, Utah, and Idaho total 3.2 billion barrels. Although this area had long been considered a potential oil resource, the vast size of the reserve was not anticipated in earlier estimates.

A similar complex fault structure roughly parallel with the Appalachian Mountains and extending from Vermont to Alabama has been designated the Eastern Overthrust Belt. Several gas wells have been drilled in Tennessee and West Virginia, and the Potential Gas Committee for the Appalachian Area has estimated reserves at 68 trillion cubic feet (Tcf). The promise of high prices for natural gas under the 1978 Natural Gas Policy Act has mitigated the financial risks and stimulated interest in exploration of this province. However, the gas is trapped in impermeable rock, and even under the most optimistic assumptions, it will take many years to develop significant production.

Outer Continental Shelf

The Atlantic continental shelf was long considered a frontier area whose exploration would result in major oil and gas discoveries. The potential of its Baltimore Canyon area was once estimated at 0.8 billion barrels of oil and 13 Tcf of gas, and oil companies paid $1.4 billion for the privilege of exploring it. Of the 23 holes that have been completed, 5 contained gas and 18 were dry. Two holes are still in progress, but the expected potential has not been realized, and the leases may be abandoned. The Georgia Embayment has also been a disappointment.

A relatively new factor affecting oil-supply planning is the environmental

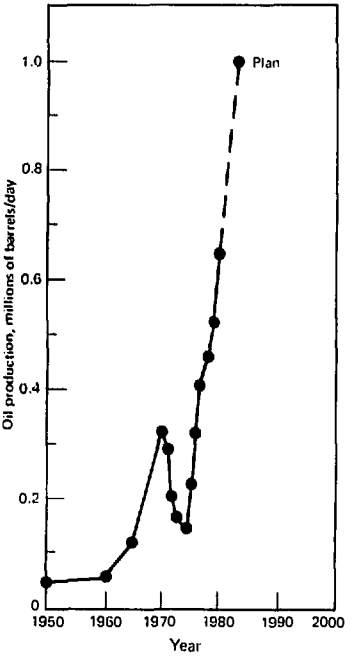


Fig. 3

Egyptian oil production, most of which is exported. The decrease between 1970 and 1975 was caused by the capture of oil fields during the Arab-Israeli war. The continuing spurt in production followed the restoration of those fields and the development of others in the Gulf of Suez.

Present exploration [in the Western Overthrust Belt] was made possible by improved seismic methods that have identified drilling targets beneath what was previously believed to be thick basement rock.

protection movement. Environmentalists can delay development for years after promising continental-shelf areas become available for exploration. For example, the Georges Bank area off Connecticut, Rhode Island, Massachusetts, New Hampshire, and Maine was opened for exploration only after long litigation with fishermen and environmentalists. An eight-year drilling moratorium imposed after the 1969 oil spill in the Santa Barbara Channel delayed until this year the production of 0.08 mb d from the Honda field, discovered in 1968. Production in the Beta field in San Pedro Bay, discovered about the same time as the Honda field, may not reach its potential of 0.05 mb d until 1986.

Production at the super-giant Prudhoe Bay field may peak in a decade or so. Unless new sources of oil are found nearby, the giant Alaskan pipeline could run dry before it is paid for. Suits by Eskimo villages and the National Wildlife Federation delayed exploratory drilling in the potentially rich area of the Beaufort Sea, north and northeast of Prudhoe Bay, until November 1980.

Policies shift unpredictably over time intervals that are short in comparison with the lead times common in the oil industry.

The Perils of Prediction

The pitfalls that lie in wait for even cautious and conservative predictors of oil supplies are most evident in the political arena. Policies shift unpredictably over time intervals that are short in comparison with the lead times common in the oil industry, and it is often impossible to decide in advance what the results of a given policy will be.

Nationalization in Canada

In 1980, the Canadian government announced a new energy policy that will levy an 8% tax on oil and gas revenues, phase out depletion allowances within two years, and acquire and transfer some oil and gas assets now held by multinational oil companies to Petro-Canada, the national oil company. The announced purpose of this policy, in addition to increasing tax revenues without burdening other industries, is to stimulate Canadian oil companies at the expense of the multinational oil companies and to increase Canadian ownership of a national resource.

In a move that probably was not foreseen by the framers of this measure, the Province of Alberta, hoping to force a reconsideration of the policy, has threatened to curtail oil production as much as 15% (it now produces 1.3 mb/d), starting with a 5% cut on March 1, 1981. It also has suspended provincial action on proposals to bring two tar-sand plants into production in the mid-1980s.

Government representatives argue that the Federal policy is in keeping with Canada's best long-range interests. They claim that even with these measures, increased oil prices by the end of the decade will provide ample cash for exploration. By discouraging the operation of multinational oil companies in Alberta, they hope to encourage Canadian frontier and offshore exploration.

In fact, the policy move frightened investors and caused many of them to dump their stock on the market, depressing its price. The affected oil companies claim that this was part of the government's strategy—to deflate the value of the companies and then buy them at bargain rates. The combination of prospective reduced cash flows and limited ownership has also caused a sudden decrease in exploratory drilling in most Canadian provinces.

The Iran-Iraq War

When conflict between Iraq and Iran broke out late in 1980, immediately depriving the world of 4.1 mb/d of oil, there was alarm at the prospect of the damage a prolonged war would do to our economy and those of our allies. However, after more than six months of fighting and the exporting of very little oil from either Iran or Iraq until recently, there has been surprisingly little dislocation in the world's economies.

The numerous reasons for the importing countries' ability to deal with this supply perturbation include strong price-driven conservation measures, increased oil shipments from Saudi Arabia, a mild world recession, and the presence of unusually large inventories (equivalent to a 110-day supply) against which the importing countries could draw.

Part of the reason for the large size of these inventories lies in the new crude-oil sales contracts negotiated by the oil-exporting countries. New contract provisions include prices that are subject to change without notice and various other measures that force the purchaser to accept large amounts of oil even when demand is falling off. The net effect of these and other changes is the smoothing of a producing country's operations, eliminating its need for peak capacity, and the forcing of its customers to enlarge their storage capacity.

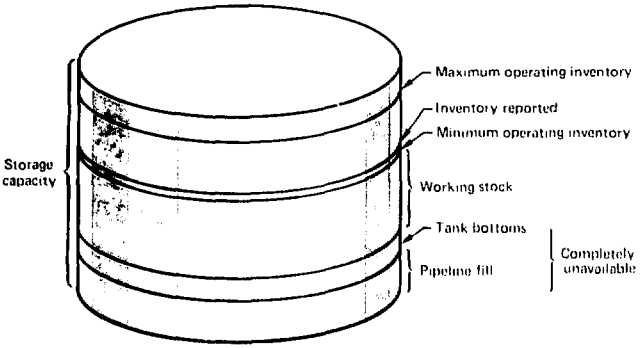


Fig. 4

Concepts and definitions associated with crude-oil storage. Inventory includes large volumes of oil that cannot be used without disrupting the system. Reported inventories include completely unavailable volumes as well as others that are essentially unavailable ("working stock").

Most industrialized nations have embarked upon government-financed and operated storage programs as a hedge against another oil embargo like that of 1973. Sweden, South Africa, the U.S., Germany, and France, among others, have built underground storage facilities that can hold many millions of barrels of oil. The Swedish (and possibly the South African) program is quite large, but other countries (including the U.S.) have so far accumulated only the equivalent of a few weeks' imports in addition to their commercial storage.

Most industrialized nations have embarked upon government-financed and -operated storage programs as a hedge against another oil embargo like that of 1973.

Only a small part of the oil in any storage system associated with refining or

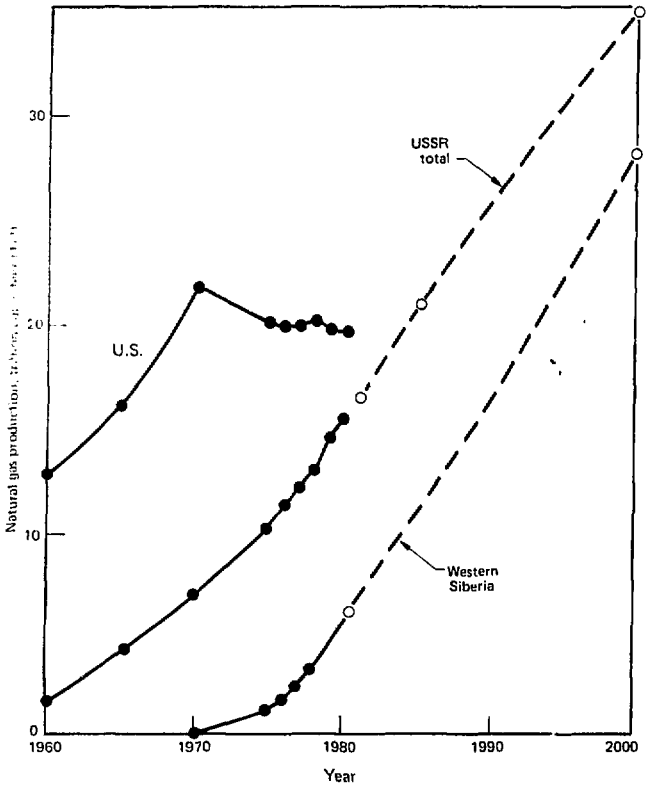


Fig. 5

Natural gas production in the U.S. and USSR. Open circles are from announced Soviet plans. According to the projection, by the year 2000, Western Siberian gas could account for 80% of the total USSR supply.

pipeline operations is immediately available (Fig. 4). Much of the system's volume is devoted to pipelines, which must remain filled in order for the system to function. Even more volume is occupied by sludge in the bottoms of tanks. Some tanks must be empty when others are full (so that there always will be a place to put oil), further reducing the actual inventory below the theoretical storage capacity. Thus, only about half of the oil inventory will be available. Usable, primary oil stocks typically comprise only about 20 to 25% of the total inventory.

Furthermore, it takes time to build large storage facilities. The only real elasticity in the whole supply system is in the seaborne link. When more oil is being shipped than can be accepted, the tankers steam slowly, saving fuel and adding the tanker volume to the storage system. In fact, many tankers that otherwise might have been scrapped have been converted instead to moored storage.

Nevertheless, the build-up of inventories can fairly be said to have reduced our vulnerability to supply disruptions. It might seem, in retrospect, that such factors could have been taken into account in assessing the potential impact of the Iran-Iraq war, but this is easier said than done. It is notoriously difficult to calculate the actions of an interactive system in which each part influences the actions of every other part.

Natural Gas from the USSR

The development of Russia's natural gas reserves, which exceed the combined reserves of all Middle Eastern countries, poses a delicate question for the countries of Western Europe. Development has been spectacular (Fig. 5) and seemingly is limited only by pipeline and compressor-station construction. Furthermore, the money Russia could earn from these resources ($3.2 billion in 1979, up 48% from 1978, about half of it coming from Western Europe) could more than compensate for the loss of revenue as Russian oil production declines over the next decades.

According to present trends, by 1990 natural gas may account for 15 to 20% of the total energy use in many Western European countries. Most of this will be imported from the USSR. The question, then, is whether it is better to be dependent on oil from the Middle East or gas from Russia.

Conclusion

Long-range forecasts are vital oil-industry planning tools because of the massive size, enormous expense, and long lead times of typical construction projects. Making such forecasts is difficult even when prices are relatively stable and the interplay of market forces is the dominant driving mechanism. Technological innovation complicates the task of the planner by unpredictably augmenting reserves, but such changes generally are welcome. However, shifting political winds and policy changes, which

often come on short notice and have unpredictable side effects, can make the planner's task a difficult one. Oil exploration and development traditionally have been gambles, but political and economic-policy factors make the risks even greater.

Key Words: economic forecasting, long range planning, oil exploration, oil-field development

Notes and References

1. L. V. Borg, *Current Development Affecting Future Availability of Oil and Gas in the Free World*, Lawrence Livermore National Laboratory Rept. UCRL-53129 (1981).
2. Sources for data in Table I: Hubbert, M. K., *U.S. Energy Resources, A Review as of 1972*, Committee Preprint, Committee on Interior and Insular Affairs, U.S. Senate, 1974, p. 19; Langston, as reported in *Oil and Gas Journal* 78(12), 29 (March 22, 1976); Miller, B., *et al.*, *USGS Circular 725*, 1975; Moody & Esser, quoted in *World Oil* 181(4), 17 (September 1975); National Academy of Sciences, *Mineral Resources and the Environment*, 1975; National Petroleum Council, *U.S. Energy Outlook: Oil and Gas Availability*, 1973; Rand Corporation, Report R-2615-I; USGS [DOI 1981, United States Geological Survey, Open File Report 81-192, 1981.

Past Titles

Listed below are articles published in recent issues of the *Energy and Technology Review*. They are grouped in terms of their relevance to the areas of responsibility of Assistant Secretaries of the U.S. Department of Energy, with a few additional categories created for research funded by other agencies.

DEFENSE PROGRAMS

International Security Affairs

Verifying the Test-Ban Treaties (June 1981)

Military Application

Brazing with a Laser Microtorch (May 1981)
A Low-Energy X-Ray Spectrometer for Pulsed-Source Diagnostics (May 1981)
Micromachining (May 1981)
Validation of the Pulsed-Power Design for FXR (May 1981)
Shock-Wave Studies: Modeling the Giant Planets (April 1981)
Light Propagation in Optical Fibers (March 1981)

ENERGY RESEARCH

Tandem-Mirror End Plugs for Future Fusion Reactors (June 1981)
TMX Experimental Results (June 1981)
Thermal Control of the MFTF Superconducting Magnet (May 1981)
Flame-Quenching in Internal Combustion Engines (April 1981)
Modeling of Solid-State Materials for Solar Cells (April 1981)
Computer simulations of ion-surface interactions (February 1981)
Improving the performance of high-speed computer networks (February 1981)
Making better scientific use of drill holes (February 1981)
Microearthquakes in the Livermore Valley (February 1981)
Physically reasonable interpolation methods (February 1981)
Underground imaging: pictures of the world below (February 1981)

LABORATORY REVIEWS

Energy and the Environment (July 1981)
LLNL 1981: Technical Horizons (July 1981)
National Defense (July 1981)
The State of the Laboratory (July 1981)
Support Programs (July 1981)
The Electronics Engineer's Design Station (May 1981)

NUCLEAR ENERGY

The Chemistry of Uranium Vapor (March 1981)

WORK FOR OTHERS

U.S. Army Research Office

Mechanisms Influencing Gun-Barrel Erosion (March 1981)