OPERATOR'S CONSOLE for spiral reader, which measures prescanned, multipronged events. Overhead mirror directs film image onto table top. At the same time, television screen (right) shows view inside small aiming circle, and CRT (left) shows digital results of scanning and measuring. Light source is a high-intensity lamp. See ref. 11.

DIGITAL COMPUTERS
in physics research

COMPUTER DISPLAY SYSTEMS

A single picture may be worth a thousand words. Physicists are learning to make better use of computer graphics in problem setting and solving.

George A. Michael

COMPUTER GRAPHICS has been defined in many ways. The clearest definition, however, is still that it is an efficient (perhaps the most efficient) technique for exchanging useful information between computer and user, primarily through pictures. We will be interested here in exploring the uses of computer graphics in physics: the techniques now used and their strengths and weaknesses.

As I searched for examples that are clearly germane to physics, one of my original criteria was that physicists should agree on the appropriateness of the examples. So I asked, and perhaps not surprisingly there was very little agreement. The trouble appeared to be at least partly due to the outmoded definition of physics being used. The definition attributed to Richard Feynman, "physics is what physicists do late at night," more comfortably fits current thoughts on the subject. I leave it to you to decide whether or not the examples described here represent some aspect of physics. They at least indicate the capabilities that computer graphics offer.

HARDWARE

What devices are used in graphics? (For a more complete discussion of the display devices available, see references 1–3. The Key Data Corp provides a periodically updated review service.[4]) We shall consider all equipment as associated either with output or with input. Output equipment is intended to produce pictures based on data that exist somehow and somewhere within the computer system. The situation is more or less reversed for input; given a picture, we wish to derive a digital representation that can be operated on by the computer. Input devices can be best considered when we consider pattern recognition.

Pen and ink

The first output device is the pen-and-ink recorder. These recorders are often the digital incremental type (more commonly found in computing centers); here only pure $(\pm X)$, $(\pm Y)$ or $(|X| = |Y|)$ moves are possible, and lines with slopes other than 0, ± 90, or ± 45 deg must be made out of combinations of the permitted moves. A typical plotter can move at 300 steps per second; if each step is 0.01 inch, the effective speed is three inches per second. Other models are capable of up to 2000 steps per second, and other step sizes ranging from 0.01 to 0.00125 in. Pen-and-ink recorders may also be of the analog type, for which distance moved is proportional to the voltage-drive signal.

We can, of course, replace the pen with some other marker, for example a lamp or a knife, and the writing surface need not be paper. Pen-and-ink

plotters are generally very accurate and have excellent reproducibility, but these qualities have been less expensive to obtain in the digital version. Another advantage is the paper size, which is quite large compared with the display size of other devices.

Microfilm

In a computer-controlled microfilm recorder a cathode-ray tube (CRT) is imaged onto film by a lens array. The film transport and the CRT control electronics are controlled by computer-generated signals. For some installations the microfilm recorder receives all the data from a tape that has been written elsewhere by a computer. Even so, a few microfilm recorders now on the market include a small computer; this method is in many cases cheaper than building special-purpose equipment and also improves flexibility. Many units also offer a "hard-copy" operation, in which the picture is produced directly onto paper.

A great deal of ingenuity has been expended to make microfilm recorders produce high-quality pictures. Although some applications that use film can make do with a viewer, printing is often required. Progress has been made in providing reliable, easy-to-use equipment to develop film and to retrieve a particular picture contained, say, in a roll or on a microfiche (70 by 105 mm) card.

Some recorders now being installed are essentially intended to replace line printers; they can help ease the output bottleneck. Computer line printers are never expected to go faster than about 4000 lines per minute (currently the fastest can run at 3000 lpm) whereas some CRT microfilm recorders are capable of over 12 000 lpm. A more real limitation is how fast the film can be moved. Framing rates range from two up to about 30 frames per second. Most recorders can accommodate either 16- or 35-mm film, and a few of the newer ones allow also for microfiche (105-mm) film widths.

Direct observation

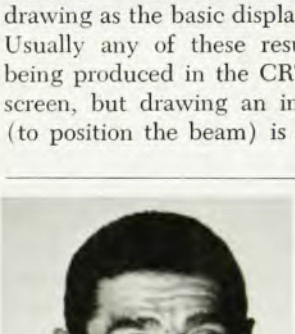
Another class of output equipment produces images that are intended to be observed directly on the CRT screen. We are interested in the more general group of direct-view CRT's that can display both text and lines. Because these devices are intended for direct viewing, the CRT is usually larger, 12 in. or more, compared to the 5- or 7-in. tube used for microfilming. There are various ways to sustain the images being presented: the display data can be stored in the computer's memory and continuously cycled to the display; the display can be equipped with its own refresh memory, or the display itself can be the self-storing variety.

Regardless of the method he choos-

es, the user must construct his pictures from the elementary functions that the display can perform. We shall simplify matters by distinguishing point plotting, character plotting and line drawing as the basic display functions. Usually any of these result in light being produced in the CRT phosphor screen, but drawing an invisible line (to position the beam) is often done, as is plotting a blank character (to separate two words). These functions are almost invariably carried out with respect to some position on the display screen, for example, drawing a line to a point (X, Y) from wherever the beam last was.

In most cases we imagine a two-dimensional cartesian lattice spanning the screen, although recently a few display systems have been defined that can operate with three-dimensional coordinates to form some sort of perspective mapping onto the display screen. Most display systems allow $2^{10} \times 2^{10}$ (addressable) points in the lattice; a few allow up to $2^{12} \times 2$.[12] The user's data, however, can span a much larger range and be mapped onto the display lattice in a variety of ways. The user also has some control over brightness and image texture (dashed lines). Times involved to execute these functions range from 2 to 60 microsec.

The direct-view CRT is the part of a bidirectional communications channel that allows the computer to show what it is doing. The other part consists of keyboards, light pen, tablets, and so on. These components help the user communicate with the computer as it executes his programs. How they are used is largely up to the user; their great value, in fact, is that they are programmable.



George A. Michael, a physicist at the Lawrence Radiation Laboratory in Livermore, California, first came to Livermore in 1953, soon after he received his bachelor's degree in physics from the University of San Francisco. Initially he did research in hydrodynamics and nuclear physics, but since 1960 most of Michael's efforts have been devoted to the development of graphic data-processing techniques and large computer systems.

COMMUNICATION

The implied dialogue between user and computer suggests that the computer might need a description of how pictures are perceived by humans. Despite all efforts, this problem is still largely unsolved. In most applications, fortunately, we can circumvent the problem of human perception. We do not tell the computer about pictures; the computer instead is treated as an infinitely obedient servant that will do exactly what we tell it to do (and, unfortunately, not necessarily what we mean it to do).

In general then a task consists of the part we believe to be precisely defined (allocated to the computer) and the remainder (retained by the human). In the sense used here, something is "precisely defined" if we can supply a computable algorithm for it. For example, although digital computers, strictly speaking, can not solve differential equations, we can get an acceptable approximation from finite-difference schemes and finite digit arithmetic. What techniques are

)MPUTER-GENERATED FILM on stratified
ear flow in a stable, viscous fluid.
e six frames show the evolution, in a
vitational field, of a finite temperature
rturbation on top of a mean
nperature gradient and shear. Computation
nsists of nonlinear equations for vorticity
d temperature. Initial perturbation
nsists of a hot spot on the left and a cold
ot on the right (top left). After a
v computational cycles, the rising warm spot
d falling cold spot induce a
ocity field; the next two frames (top
nter and right) show the streamfunction and
rticity appropriate to this field,
d the remaining frames show composites of
e three fields. (The film itself is in
lor, so that each field is clearly
tinguishable.) After a time, hot and cold
ots exchange places (center); after
other period they switch again (right).
e vorticity–streamfunction correlation
en in the three composite frames tells
e experimenter that the nonlinear advection
m in the vorticity equation may be
nored. See refs. 12 and 13.

CORRELATION MAP OF AN "OH" SOUND made by experimenter Jack Oliver. A microphone converts the sound into an electrical signal that varies with time. After an analog-to-digital converter samples the signal 6000 times per second, the digitized output is truncated to nine bits and stored. The display corresponds to plotting the points $(\alpha S_i, \beta S_{i+1})$, $i = 1,2 \ldots N$; here the S_i form a sample set of N observations, and α and β are scale factors. At the same time that the set is plotted, an audio signal is derived from the subset of y (βS_{i+1}) coordinates and is fed to the speakers. Although the visual shape suggests a possible new way to recognize speech, there have been as yet no positive results from these experiments.

used for these "precisely defined" segments?

Lists

In one common method, the user must somehow provide a digital description (usually in list form) for each item that might be pictured. This description effectively tells how to construct the image. A list is usually made up of a *header* that contains relational information and a *body* that contains data. For graphics, the actual image data or pointers to them are typically put into the list body, and the header contains such things as pointers to other lists, membership designators and other processing directions.

The particular ordering of the image data, although not necessarily unique, is usually important because it does imply the sequence in which the figure is to be drawn. A quadrilateral, for example, can be described as five number pairs, the last being a repeat of the first. The number pairs represent vertices A, B, C, D, A, and the actual drawing operation is: "move invisibly to A, draw to B, then to C, then to D, then to A." For efficient processing it usually is necessary to ensure that each list is structurally analogous to all the other lists.

If the method is properly designed, one member can represent several instances of use. Although list-processing techniques were developed largely in connection with studies in formal logic, theorem proving, artificial intelligence and certain types of language processing, they also turn out to be extremely useful in certain computer-graphics applications. (For details of several list-processing languages, see references 5–7.)

Not all computer-graphics applications use these list-processing techniques. In some programs the displayable data are produced by computation. A user might need or wish to do this for a variety of reasons: to generate the loci of a family of curves, for example, or to see certain characteristics as a given parameter is changed. Or he may not have enough storage space to hold all the quantities to be displayed, but yet have enough room and time to compute them when needed.

A very common situation arises

when display capability is added to an already existing program. Here the user might employ subroutines that can be called up from strategic places in his program. What he produces is mostly up to his own imagination. Graphs can be made easily, and pictorial rendering can also be done—in some cases even animated.

Pattern recognition

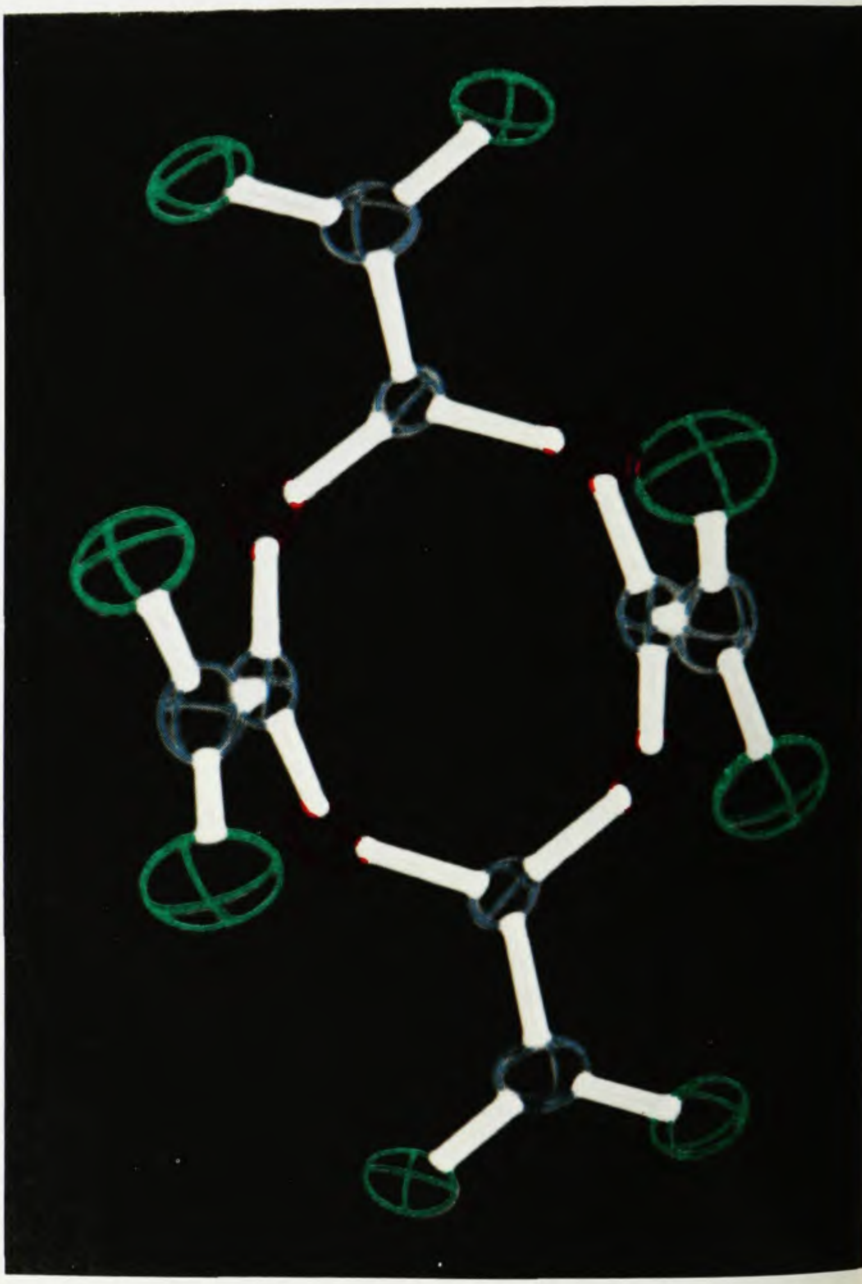
The main criterion of the type of method we have been discussing is that either the user or his program supplies descriptions of picture parts whose ordering implies the sequence in which they are to be displayed. An alternative method is used mostly for the part of computer graphics known variously as "pattern recognition," visual data processing" or "picture processing." The user must be able to define real pictures precisely. In contrast to the first method, in which the user effectively provides a stylized rendering of the objects to be displayed, here the information is either on or must be derived from data that already exist in a photograph and nowhere else. This is the case for bubble-chamber photographs, star and satellite pictures, oscillographs, seismograms, electromicrographics and electroencephalograms, for example. The raw material for the grandest adventures is blended with all manner of unwanted photographic artifacts—in a word, noise. And the coordinate data that represent a given image are not nearly as concise as are the simple lists of lines.

To get an idea of the extreme data storage capacity this method might demand, consider what it means to obtain a "precisely defined" digital equivalent of an ordinary photograph. Because we have no theory of visual perception (pattern recognition) to guide us, we must abandon the attempt to organize picture content according to the images "known" to be there. If, however, we know (up to the limit of resolution) a numerical value for the photographic density at every point, we then have, in principle, all the data it is possible to get. Given the coordinates of a point, we can obtain the value of the photographic density at the specified point.

Scanning and tracking

One ordering method is to arrange the values in the same sequence that would occur if the points were interrogated by a beam programmed to move as does a television scan. This technique is well suited to the way computers operate, but results in enormous amounts of data. Many films used in science applications can routinely resolve 500 lines per mm, and variations in the optical density span at least four orders of magnitude. A 25-cm square plate thus represents over 10^{10} resolution elements. If the densities at each point are obtained to only three significant figures, we would need of the order of 10^{11} bits of computer storage, roughly equal to the number of core memory bits in the entire US! The digitizing process must either be done at much lower resolutions or for only a tiny piece at a time. For some applications it is possible to combine both techniques: a coarse scan to provide a general view followed, if appropriate, by higher-resolution examination of selected areas.

For other applications, the user must know what he wants from the picture—the locus of a series of bubble

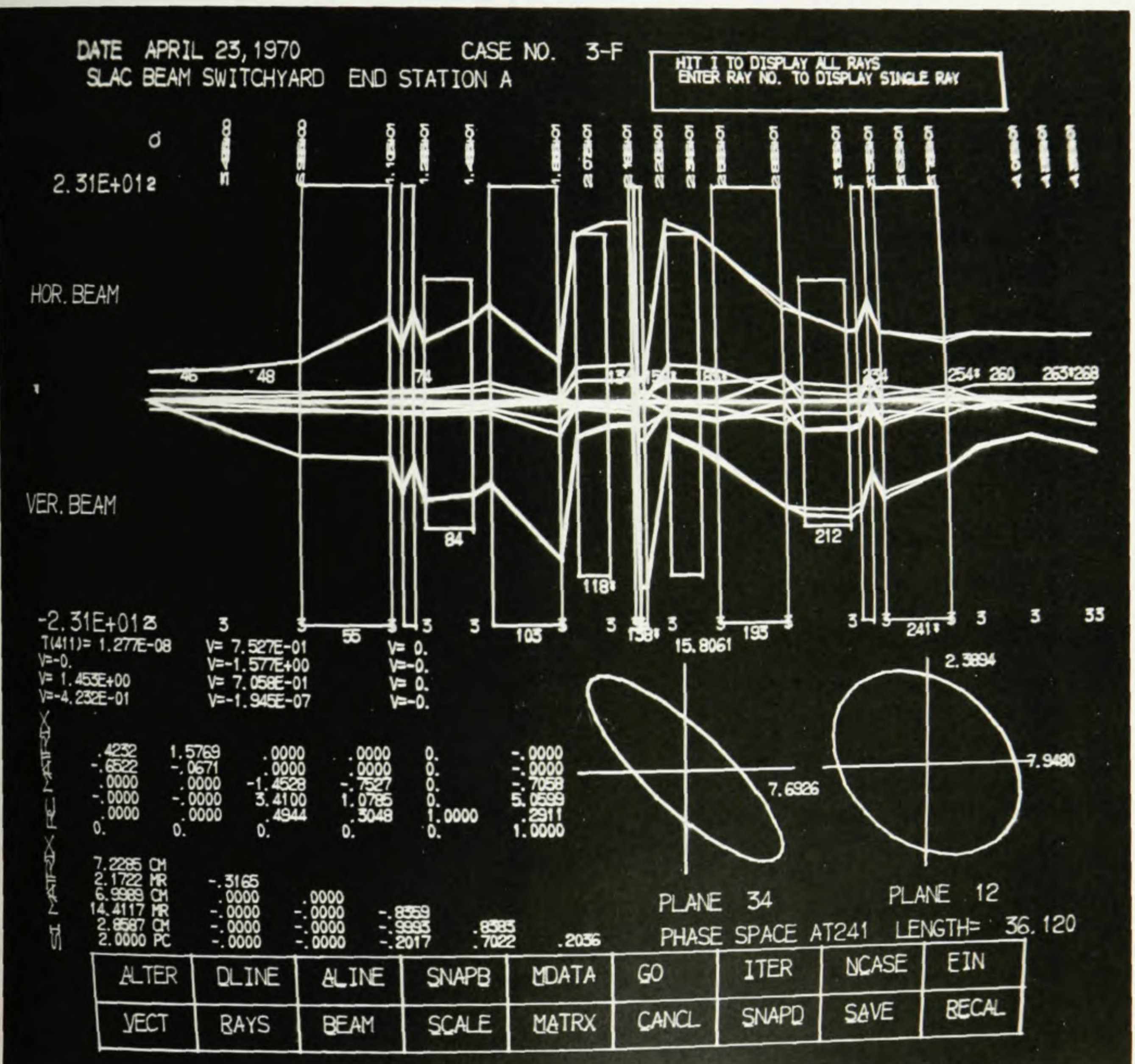


CRYSTAL-STRUCTURE PLOTTING PROGRAM. Crystallographer plots two-dimensional data and sees stereoscopic three-dimensional views of molecules. Compound in this CRT photograph is octahydro-1,3,5,7-tetranitro-1,3,5,7-tetrazocine; green is oxygen, blue is nitrogen and red is $-CH_2$. Ellipsoids are probability surfaces determined by thermal motions of the atoms, which are very hard to represent with "ball-and-stick" models. In the color film from which this print was made, the figure rotates so that the shape can be seen from various angles. See refs. 15 and 16.

images, an oscilloscope trace, a density profile at some given location—so that only the explicit data need be extracted. For this method, called "tracking," the data themselves can be explicitly ordered, one bubble following another, whereas for scanning, data ordering depends on the television-like format. Tracking permits the preservation of relational and contextual information, whereas in scanning this information, if needed, must be extracted through subsequent digital processing of the collected data.

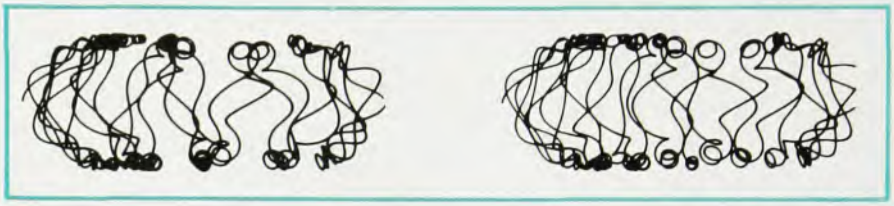
The equipment needed to digitize consists of a light source, a document holder, appropriate lenses, light detectors and various electronic units for power and signal controls. Historically these devices were known as "flying spot scanners." Most machines used in high-energy physics research programs allow at least some program control of beam motion.[8] The light source is either a CRT or a high-intensity lamp. (See figure on page 31.)

A typical use of scanning is image enhancement, sometimes called "picture filtering." The optical density for each picture element in a given set is obtained (read), transformed according to some algorithm and used either to produce a new picture[9] or as input to a series of pattern-recognition algorithms.[10] If the user is able to use the tracking method, then in general fewer data points are obtained, although many more points may be interrogated on the film. The actual



INTERACTION OF CHARGED-PARTICLE BEAM with controlling magnets can be evaluated from a display of transformation matrix and beam size along the flight path. Display helps the user to select the most useful experimental parameters. Any of the magnet or beam parameters may be altered and the effect of the alteration determined. The effect of energy spread on the beam, for example, may be readily evaluated either by altering ΔE in the calculations or by simultaneously displaying vectors that comprise a second beam of different energy. See ref. 14.

FOURIER EXPANSION OF A STEP FUNCTION (in a pair of stereoscopic views); each of the four approximations includes one additional term. The Gibbs effect, which predicts an overshoot no matter how many terms are included, is already obvious, and the experimenter can see (by calling for more terms) that the fit at the edges of the step function will never improve by very much. See ref. 3.



EARTH'S MAGNETIC FIELD approximated by a bar magnet. Stereoscopic view (N. Hardy) shows three-dimensional trajectories of trapped charged particles, which "rattle" from pole to pole. Varying the angle of injection produces different results.

reading process is usually supervised by a human operator.[8,9,11]

APPLICATIONS

Activities in physics are sometimes classified as either theoretical or applied. In relation to computer usage, there is a more suggestive division: modeling and data processing. Modeling, that is describing a physical system with mathematics, is of course very old, but until computers were available it was usually quite difficult to perform the calculations. There are two types of modeling that computers serve well. First, if we already have an accepted system of equations that describe a class of physical phenomena, we can perform computations to examine various logical consequences. That is, we use the model for prediction and verification. Alternatively the computer allows us to choose from among several mathematical formulations in attempts to develop a suitable theory for a given set of phenomena, or to establish the values of a set of parameters. Either way the model is expressed as a computer procedure, and we can contrive to obtain a variety of pictures from it—pictures that can help verify the physics, stimulate intuition and suggest new ideas.

For the other kind of computer use, processing experimental data, we still must do the same kinds of things that were always done, but it is, or can be, vastly easier if a computer is used. Because the data must be in a form that the computer can read, more and more experimental setups are being designed to include a computer for both control and data logging.[8,9] Addition of a display console permits the experimenter to interact with the experiment, often while it is going on. This interaction can mean that better data are obtained, and perhaps fewer experimental points need be measured. If, as in the case for high-energy physics, the experiment produces data too fast for the computer to record, or if more data is produced than can be stored, the data may first be captured on film and a film reader used to get the data into computer-readable form.

In general, all the examples illustrated here exhibit, in varying degrees, both modeling and data-processing characteristics. Illustrations are either taken from a computer-generated motion picture or represent an instant in a display associated with an interactive graphics application. Neither the seminal effect of actual interaction with a computer display, nor the chance of a suddenly new intuition while watching a computer-generated motion picture can be communicated through the illustrations, but they at least hint at these effects.

References

1. S. Davis, *Computer Data Displays*, Prentice-Hall, Englewood Cliffs, N. J. (1969).
2. *Recent Advances in Display Media*, NASA, Washington, D. C. NASA SP-159 (1967).
3. G. Michael, R. Cralle, "A Survey of Graphic Data Processing Equipment for Computers," in *Computer Oriented Circuit Design* (F. Kuo, W. Magnuson, eds.), Prentice–Hall, Englewood Cliffs, N. J. (1969).
4. Computer Display Review, (J. Martell, ed.) Key-Data Corp., Watertown, Mass. (1970).
5. K. Bertran, G. Michael, "Languages for Graphic Data Processing," in *Computer Oriented Circuit Design* (F. Kuo, W. Magnuson, eds.) Prentice–Hall, Englewood Cliffs, N. J. (1969).
6. *Programming Systems and Languages*, (S. Rosen, ed.), McGraw–Hill, New York (1967).
7. D. Bobrow, B. Raphael, Comm. A. C. M. **7**, no. 4, (1964).
8. *Methods of Computational Physics, Vol. 5, Nuclear Particle Kinematics*, (B. Alder, S. Fernbach, M. Rotenberg, eds.) Academic, New York (1966).
9. *Pictorial Pattern Recognition*, (G. Cheng, R. Ledley, et al, eds.) Thompson Book Co., Washington, D. C. (1968).
10. *Optical and Electro-Optical Information Processing* (J. Tippett, D. Berkowitz, et al, eds.), MIT Press, Cambridge, Mass. (1965).
11. G. Lynch, *Remarks on the Use of the Spiral Reader*, UCRL-17328, Feb. (1967).
12. W. P. Crowley, G. M. Corcos, *Numerical Calculations of Viscous, Incompressible Stratified Shear Flows*, UCRL-50538, Nov. (1968).
13. H. E. Schwember, *The Interaction of Buoyancy, Shear and Diffusion in Stratified Flows*, PhD Thesis, Univ. of Calif. (1969).
14. J. S. Colonias, A. C. Paul, *On Line Calculations of Beam Transport Systems to First and Second Order*, UCRL-19414, April (1970).
15. Q. Johnson, G. McMillan, G. Smith, *Stereo Movies of Crystal Structures*, UCRL-14717 Abstract, Feb. (1966).
16. C. K. Johnson, *ORTEP: A Fortran Thermal-Ellipsoid Plot Program for Crystal Structure Illustrations*, ORNL-3794, June (1965). □