NEW INFORMATION PROGRAM FOR AIP

How do you cope with the ever-increasing flood of literature? A new computer-assisted system will offer new and better ways of obtaining physics information. We seek your opinions.

ARTHUR HERSCHMAN, FRANZ L. ALT and H. WILLIAM KOCH

THE AMERICAN INSTITUTE OF PHYSICS, with support from the National Science Foundation, is currently engaged in a major effort to develop and implement a computer-assisted "National Information System for Physics." Designed by physicists, for physicists, the new system is scheduled to begin pilot operations early next year. We at AIP believe the system to be urgently needed, but how do you, the physicists, feel about it?

This article presents a description of the main features of the new system so that you can form an opinion on its merits and its potential usefulness. After you have read the article, we hope to hear from you on this important question: Do you feel there is need for this program and is it aimed in the right direction?

AIP responsibility

Because we believe that there is a need for a physics-information system and that AIP is the logical place for its development, we have assumed the responsibility and undertaken the formulation and development of a new system.

AIP was founded in 1931 as a federation of leading societies in physics to serve those needs of the physics community that could best be fulfilled by the societies jointly. It presently has seven member societies, whose 47 000 members are also institute members, plus 19 affiliated societies with an interest in physics, 150 corporate associates and a Society of Physics Students. The institute services for this sizable community run the gamut from publicizing physics and physicists, strengthening educational programs, documenting the history and development of physics and representing physics nationally and internationally—to the largest single publishing effort for physics in the world. AIP publishes 16 archival journals, comprising 25% of the world's articles in physics and translates 13 Russian journals, for an additional 10%.

As a natural extension of its responsibility and in accord with its mandate to engage in activities "for the advancement and diffusion of the knowledge of the science of physics," AIP, with support from NSF, has been actively planning further information services since mid 1966.[1] These plans led to our program for the design and development of a national information system for physics. At the end of 1967, a new division was organized within AIP to handle the project.[2]

The division has the assistance of a 15-member advisory committee, which was appointed by AIP member societies, of about 100 physicist-respondents selected by the advisory committee and of liaison members from other interested groups, both from related scientific societies (chemistry, mathematics and engineering) and from interested government agencies. The results of this effort, a national physics-information system, will be ready for implementation during 1970. (A document describing the proposed system was recently presented to NSF in support of a request for funding the pilot operations.)[3]

Why a new system?

A new system is needed to cope with the exponential growth of physics literature, which has been doubling about every seven and a half years. It

Arthur Herschman (left) has been director of the AIP information division since its inception in 1967. A theoretical physicist, who received his PhD from Yale University in 1954, Herschman was formerly coeditor of *The Physical Review*.

Before becoming AIP director in 1966, H. William Koch (center) was chief of the radiation-physics division at the National Bureau of Standards. Koch joined NBS in 1949, after receiving his PhD from the University of Illinois, and worked in the high-energy-radiation section until becoming division chief.

Franz L. Alt (right), who took his PhD in mathematics at the University of Vienna, became deputy director of the information division after 19 years with the National Bureau of Standards. At NBS he was assistant chief of the applied-mathematics division and, later, area manager for information systems, design and research.

is not that physicists are writing more, but that more physicists are writing—more physicists in every speciality. In 1968 alone, over 50 000 research papers were published in more than 500 journals. Finding information in the traditional way, by scanning journals and through formal and informal talks at meetings, is no longer practical. Although one may still keep up with new developments of immediate interest, it is almost impossible for any one physicist to keep abreast of bordering areas and related specialities with which he should be familiar.

The tendency of the present procedures to be designed for authors' convenience has also aggravated the problem. Authors, not readers, determine when, where and how the information is presented. As a result, papers on any given subject are dispersed over many journals, and a single journal may contain, side by side, papers on widely different subjects. The reader is left to cope with the flood as best he can, which all too often results in information coming to his attention too late for his need.

What is clearly required is a better way to organize and manage the information so it can be routed more accurately and efficiently from author to reader.

Computerized file

The only feasible way to organize and manage a collection this large is by computer—to have a computer record of each new paper that allows a file to be organized and searched on a current basis according to physicists' interests.

As presently conceived, the file would initially contain records for about one half of the world's physics-journal articles, but would be expanded to cover almost all journal literature, as well as nonjournal material, in the not-too-distant future.

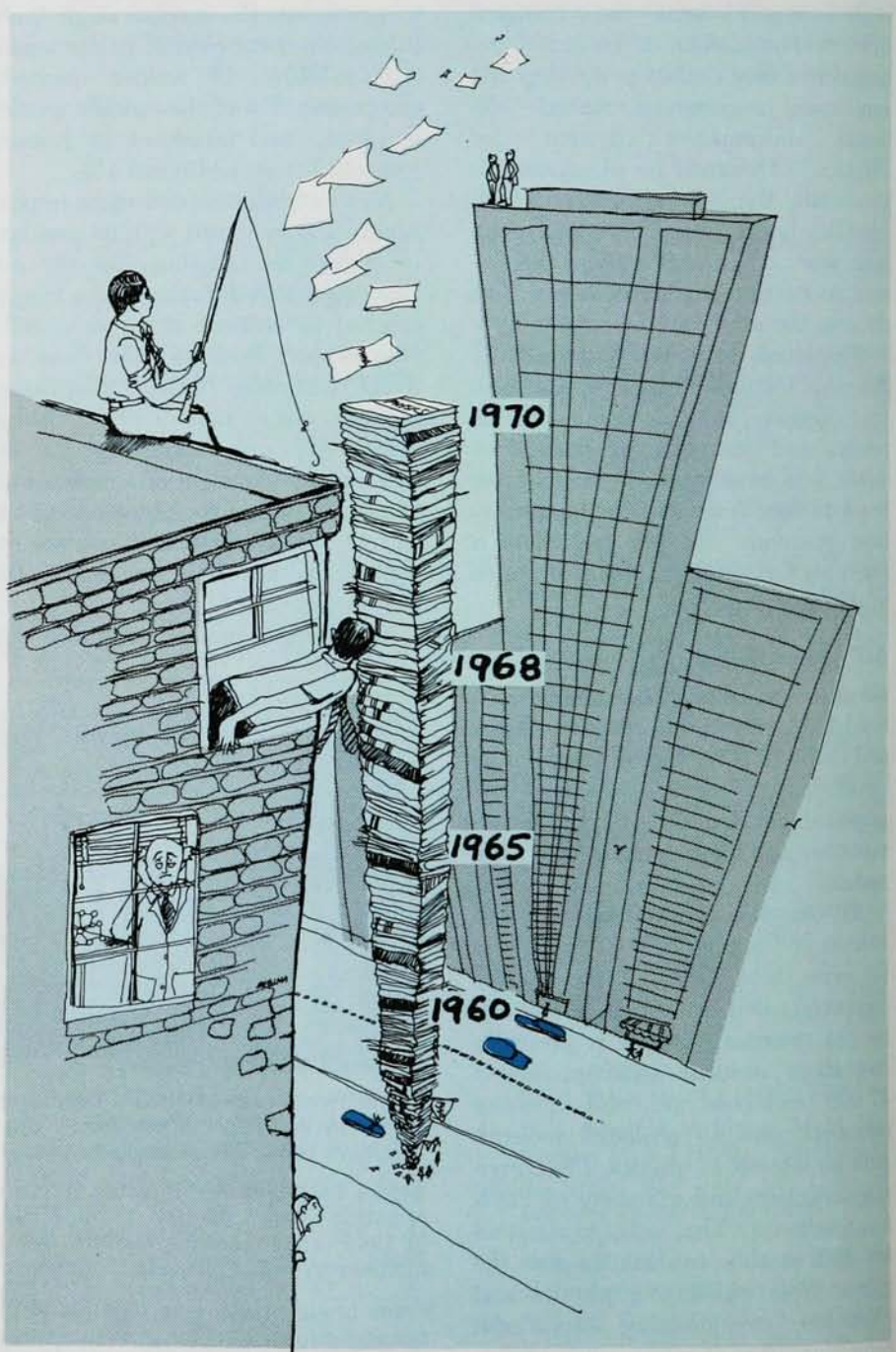
Every month, for AIP-published journals during the prepublication cycle and for other journals as they are received, the information-division staff will prepare, for each new paper, a "record" that contains basic information about the paper: author, journal, title, abstract, citations (that is, references to other literature), a list of "key words" and a special "AIP classification number."

These records are then transcribed onto magnetic tapes. Thus the cumulative file of all such records constitutes, in effect, a machine-searchable "physics almanac" that can be queried for a multitude of purposes and produces a variety of services. Specially formatted printed versions of all or part of the file can be widely distributed for ready reference. The file itself could answer specific questions both at AIP and at suitably equipped subscribing institutions.

Each item in the article record represents a "handle" that can retrieve the complete record. Thus one can ask for all articles published in a given journal or year or by a particular author or institution; papers that contain certain specific words in their title or abstract or that cite another paper or have a number of citations in common with a given paper; and finally, papers about a particular kind of physics.

Classification scheme

The classification procedure that the system as a whole will use was developed by AIP in coöperation with outside physicists specializing in various branches. It is a procedure for writing a formatted statement of what, objectively, a paper is about. The box on page 31 is an example and shows how the classification number is constructed using "verb expressions," represented as integers, followed by "nouns," represented as decimal num-



"THE READER IS LEFT TO COPE with the flood as best he can . . ."

bers. Typical verb expressions are: "the subject of primary interest is . . ." and "the method used is . . ." and "the host or environment is . . . ."

The more digits in a numerical noun, the more specific its meaning. For example ".2" means *particle*, ".28" means *hadron*, ".282" means *baryon*, ".2821" means *nucleon*, ".28211" means *proton*.

This particular way of "spelling" nouns has the advantage of exhibiting the word roots. Suppose, for instance, you want papers on hadrons. All nouns beginning with the digits ".28" belong to the class hadron. Knowing this, a request can easily be formulated. The same request, in clear language, would require a specification of all words included in the class hadron (meson, pion, kaon, baryon, hyperon and nucleon). In the example (right) the title is less explicit, from the viewpoint of information retrieval, than is the classification number.

CLASSIFICATION EXAMPLE

For the paper "Evidence of Quarks in Air-Shower Cores"
0.1; 1.271; 2.9534; 4.24; 6.29

0	The document type is . . .
0.1	experimental;
1	The subject of primary interest is . . .
1.2	particle physics,
1.27	more precisely, a particle property,
1.271	specifically, its existence;
2	The method used is . . .
2.9	a technique,
2.95	more precisely, a particle technique,
2.953	still more precisely, a detection technique,
2.9534	specifically, track visualization;
4	The entity of primary interest is . . .
4.2	a particle,
4.24	more precisely, a hypothetical particle;
6	The host or environment is . . .
6.2	particles,
6.29	more precisely, cosmic rays.

What the file will do

One of the principles underlying the design of the system was that it should evolve gradually; not only will its coverage of physics literature be increased step by step, but also its services will become more sophisticated in stages. Thus we can improve the system as we go along.

The services that will be offered during 1970 and early 1971 are all straightforward products of the information file:

• *Current Physics Titles* (CPT), a current-awareness journal, initially in four sections that probably will be: particle, field and nuclear physics; atomic, molecular, chemical, plasma and fluid physics; solid-state physics; and optics, acoustics, astrophysics and geophysics. We expect the sections will be published every other week, with each section representing a printout of the accumulated records since the previous issue. The records will be arranged, under a new system of headings, in as many places as physicists would expect to find them. The journal will be produced through computer-controlled photocomposition, so that it will be of high typographic quality.

• A series of specialized bibliographies in several of the narrower fields of physics (updated periodically) as well as indexes for the primary journals published by AIP

• *Searchable Physics Information Notices* (SPIN), a magnetic-tape service that will allow organizations with adequate computer facilities to establish their own current file of physics information. The tapes will be issued monthly and will contain the records accumulated since the previous issue. The subscribing institution could use its own search programs or specifically designed AIP programs.

At a later date, the system will offer:

• File searches based on requests. This service would be of particular value to scientists writing reviews or data compilations. Considering the importance of this activity in evaluating and distilling the literature into a more meaningful and digestible form, additional means for encouraging the production of such articles are also being planned.

• Lists of articles tailored to the needs of groups of physicists working in specialities and who do not have local facilities for using the magnetic-tape (SPIN) service, as well as procedures for subdividing journals into packages that would better suit the needs of smaller interest groups.

• Microform copies of the primary articles, as a backup to CPT and SPIN.

Long-range prospects

We expect to improve the system on a continuing basis, rendering services as effectively, and as inexpensively, as possible. As a long-range prospect, we hope to offer a centralized service, with decentralized satellites, that would cope with all the information needs of the physics community as well as those of the broader national scientific and technical community for physics information.

It would offer reference services and would obtain copies of hard-to-get material and refer questions it can not answer to sources that could. The system would also afford direct on-line access to the computer file from remote-access terminals in physics departments and other institutions. The centralized service also would have facilities for "scholars in residence," to supply clerical and reference aid for review writers.

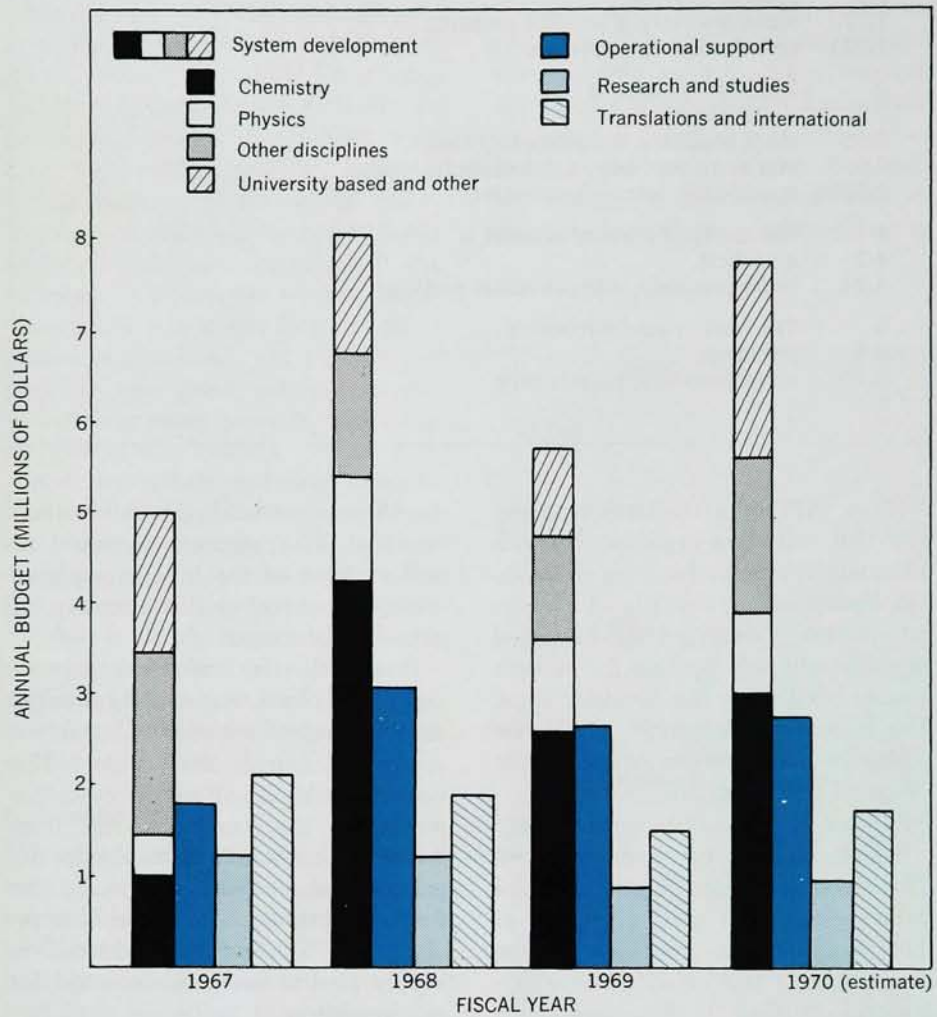
Such a centralized facility would be linked into a network of "information centers" at various institutions and of similar facilities for other disciplines and for physics in other countries. This organization, with its information file and its broad spectrum of services, linked into a network of other information centers and services would constitute the "National Information System for Physics."

Value and cost

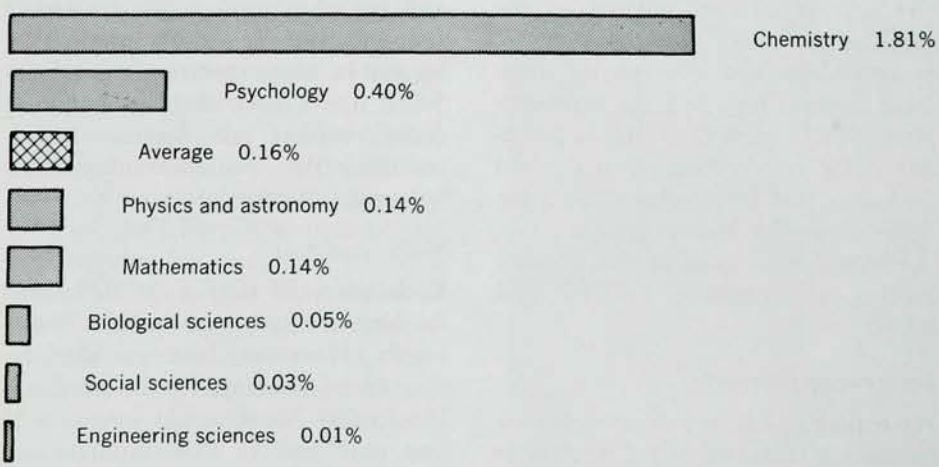
Each potential user must determine for himself what a service like this is worth. How many hours per week do you spend looking for information? How many hours would you save if you only had to look through one booklet, a short list or a response on your computer terminal? How much time would you save if the article was in one small collection or a numbered

entry on a reel of microfilm that your librarian or secretary could copy?

These questions raise a number of imponderables. To put them into better perspective, consider that each published article represents about $60 000 worth of research investment; costs about $500 to be published and will cost about $15 to process and enter into the proposed physics-information file. The distribution price for listing that article in CPT will be only a fraction of a cent per copy. If, say, one out of a hundred articles is interesting to you and one out of a thousand important, would it be worthwhile for you to have it pinpointed? Similar considerations apply to the other services. Is a system that could accomplish these things worth the cost?



COST OF INFORMATION SYSTEMS. Chart compares NSF support of information programs by type and discipline during 1967–70. **—FIG. 1**



INFORMATION VS RESEARCH COST. Chart depicts ratio of NSF support for development of information systems to all federal support of research in the fiscal year 1968. The high percentage for chemistry is partly because most chemistry research is privately financed. **—FIG. 2**

The initial cost of development and pilot operation is being funded by the NSF, as part of its nationwide support of information services in scientific disciplines. Figures 1 and 2 show NSF expenditures for these purposes, both in absolute magnitude and in relation to total research support. Is it in the national interest to have the NSF support these programs? In 1968 NSF spent about $14 million for information activities in various scientific disciplines—less than 10% of the total was for physics. The improvement of efficiency in physics research and development activities is clearly in the national interest. The saving of a fraction of an hour per week by each of the 30 000 physicists in the National Register of Scientific and Technical Personnel, not to mention all the other users of physics information, would more than make up for all of the costs borne by the NSF.

In the future many of the operating services are expected to be self-supporting after the requested funding period ends in 1973. Some of the newer services would still need subsidies, and funds would still be required for further development of longer-ranged projects. The rate of NSF support, however, would probably decrease, and the additional cost would be offset by the greater values of the ultimate system.

A question

This program has been endorsed by the AIP governing board, which represents the member societies. But considering the magnitude of this undertaking, we would like the additional opinions of individual physicists: What do you think of our proposed system? Please write us and give us the benefits of your views and ideas on this matter.

References

1. V. Z. Williams, E. Hutchisson, H. C. Wolfe, PHYSICS TODAY, **19**, no. 1, 45 (1966).
2. H. W. Koch, PHYSICS TODAY, **21**, no. 4, 41 (1968).
3. *A Program for a National Information System for Physics*, American Institute of Physics, publication no. 1D69R (August 1969). □